

1 CGGGCGCGGC GCGGCGGCG GTGACAGCGG CGCCCGCGCC TCCCCGCGCG  
 51 TAGGTGTGCG GCGCGCTCCT GCGAGGACG GAGCGAGCAG ATCTCGCGTG  
 101 CGCTCGCCGC CCGGCGCAGC CCAGCCCGGC CCCCCTGG CGCCGCGAGC  
 151 CGAGGTGTCT CCCGCGCCCG CGCCCGTGTG GCCGCGGTGC CCGCGAGCGG  
 201 GAGCCGGAGT CGCCCGCGCC CGAGCGCAGC CGAGCGCAGC CCGAGCCCGT  
 251 CCGCCGCCGC CATGGCCACC ACGGTGACCT GCACCCGCTT CACCGACGAG  
 301 TACCAGCTCT ACGAGGATAT TGGCAAGGGG GCTTTCTCTG TGGTCCGACG  
 351 CTGTGTCAAG CTCTGCACCG GCCATGAGTA TGCAGCCAAG ATCATCAACA  
 401 CCAAGAAGCT GTCAGCCAGA GATCACCAGA AGCTGGAGAG AGAGGCTCGG  
 451 ATCTGCCGCC TTCTGAAGCA TTCCAACATC GTGCGTCTCC ACGACAGCAT  
 501 CTCCGAGGAG GGCTTCCACT ACCTGGTCTT CGATCTGGTC ACTGGTGGGG  
 551 AGCTCTTTGA AGACATTGTG GCGAGAGAGT ACTACAGCGA GGCTGATGCC  
 601 AGTCACTGTA TCCAGCAGAT CCTGGAGGCC GTTCTCCATT GTCACCAAT  
 651 GGGGGTCGTC CACAGAGACC TCAAGCCGGA GAACCTGCTT CTGGCCAGCA  
 701 AGTGCAAAGG GGCTGCAGTG AAGCTGGCAG ACTTCGGCCT AGCTATCGAG  
 751 GTGCAGGGGG ACCAGCAGGC ATGGTTTGGT TTCGCTGGCA CACCAGGCTA  
 801 CTGTCCCCCT GAGGTCCTTC GCAAAGAGGC GTATGGCAAG CCTGTGGACA  
 851 TCTGGGCATG TGGGGTGATC CTGTACATCC TGCTCGTGGG CTACCCACCC  
 901 TTCTGGGACG AGGACCAGCA CAAGCTGTAC CAGCAGATCA AGGCTGGTGC  
 951 CTATGACTTC CCGTCCCCCTG AGTGGGACAC CGTCACTCCT GAAGCCAAAA  
 1001 ACCTCATCAA CCAGATGCTG ACCATCAACC CTGCCAAGCG CATCACAGCC  
 1051 CATGAGGCC TGAAGCACC GTGGGTCTGC CAACGCTCCA CGGTAGCATC  
 1101 CATGATGCAC AGACAGGAGA CTGTGGAGTG TCTGAAAAAG TTCAATGCCA  
 1151 GGAGAAAGCT CAAGGGAGCC ATCCTCACC AATGCTGGC CACACGGAAT  
 1201 TTCTCAGTGG GCAGACAGAC CACCGCTCCG GCCACAATGT CCACCGCGGC  
 1251 CTCCGGCACC ACCATGGGGG TGGTGGAAAC AGCCAAGAGT TACTCAACA  
 1301 AGAAAGCAGA TGGAGTCAAG CCCAGACGA ATAGCACCAA AACAGTGCA  
 1351 GCCGCCACCA GCCCCAAAGG GACGCTTCTT CCTGCCGCC TGGAGCCTCA  
 1401 AACCACCGTC ATCCATAACC CAGTGGACGG GATTAAGGAG TCTTCTGACA  
 1451 GTGCCAATAC CACCATAGAG GATGAAGACG CTAAAGCCCG GAAGCAGGAG  
 1501 ATCATTAAGA CCACGGAGCA GCTCATCGAG GCCGTCAACA ACGGTGACTT  
 1551 TGAGGCTTAC GCATTCTACT TCGAGAACCT GCTGGCCAAG AACAGCAAGC  
 1601 CGATCCACAC GACCATCCTG AACCACACG TGCACGTCAT TGGAGAGGAT  
 1651 GCCGCTGCA TCGCTTACAT CCGGCTCACG CAGTACATTG ACGGGCAGGG  
 1701 CCGGCCCCGC ACCAGCCAGT CTGAGGAGAC CCGCGTGTGG CACCGCCGCG  
 1751 ACGGCAAGTG GCAGAACGTG CACTTCCACT GCTCGGGCGC GCCTGTGGCC  
 1801 CCGCTGCAGT GAAGCCAAGG GAGGGGCACA GAATGGGGAA CAGGACACAG  
 1851 GATCCTAAAC TCCAAGGGGA CTGTCCACCG ATGAACACTC AGAGTGGACA  
 1901 CCATCTTCCG TCCACGCTGT GCCCAGGACA GCTGTCCCA TCCATGAACA  
 1951 CAGGGTAAAC ATCTGCCGGG CTCCGCACCA GTGGCTCCCT GGGCCATGGG  
 2001 ACAGCGGCAG GGCTCACCAC GGACAGCAGC TGGCCCAGCA GCCGGCCACC  
 2051 CTGGCGTCCT GGGGCTCCT CCCCTCCTCT CCCTCTCACC TTGTACCTC  
 2101 CACGGAGCTG CTTGTCTGGG ATAATTTGGG GATTTTTTTT TCTGGGGGAT  
 2151 AATTCTTTTG CATGACCCCT AAAGAGCAAG CCACACCGGT CTGCTAGCTA  
 2201 GGTGTCCGCG GTGTGGTG (SEQ ID NO:1)

**FEATURES:**

5'UTR: 1-261  
 Start Codon: 262  
 Stop Codon: 1810  
 3'UTR: 1813

# Homologous proteins:

## Top 10 BLAST Hits

			Score	E
CRA 18000005245285	/altid=gi 5326757 /def=gb AAD42035.1 AF07880...		1047	0.0
CRA 18000005199792	/altid=gi 10835006 /def=ref NP_001211.1  cal...		1044	0.0
CRA 18000004938668	/altid=gi 6671660 /def=ref NP_031621.1  calc...		1039	0.0
CRA 18000004937301	/altid=gi 11120682 /def=ref NP_068507.1  Ca+...		1038	0.0
CRA 18000005245287	/altid=gi 5326762 /def=gb AAD42037.1 AF08192...		1001	0.0
CRA 18000005171302	/altid=gi 3668373 /def=gb AAC79460.1  (AF085...		999	0.0
CRA 1000737074531	/altid=gi 6688228 /def=emb CAB65122.1  (AJ252...		986	0.0
CRA 18000005245288	/altid=gi 5326764 /def=gb AAD42038.1 AF08341...		986	0.0
CRA 18000004964693	/altid=gi 466360 /def=gb AAA81938.1  (U06636...		982	0.0
CRA 18000005199791	/altid=gi 4139268 /def=gb AAD03743.1  (AF112...		982	0.0

## BLAST dbEST hits:

	Score	E
gi 12801212 /dataset=dbest /taxon=960...	1675	0.0
gi 12868201 /dataset=dbest /taxon=960...	1453	0.0
gi 2053138 /dataset=dbest /taxon=9606 ...	1247	0.0
gi 10213950 /dataset=dbest /taxon=96...	1243	0.0
gi 9324431 /dataset=dbest /taxon=960...	1233	0.0
gi 12921378 /dataset=dbest /taxon=960...	910	0.0

## EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

From BLAST dbEST hits:

gi|12801212 Fetal brain  
gi|12868201 Fetal brain  
gi|2053138 Testis  
gi|10213950 Lung small cell carcinoma  
gi|9324431 uterus endometrium adenocarcinoma cell libe  
gi|12921378 Fetal brain

Tissue expression from PCR-based tissue screening panels:

hippocampus

1 MATTVTCTRF TDEYQLYEDI GKGAFSVVR CVKLCTGHEY AAKIINTKKL  
 51 SARDHQKLER EARICRLKH SNIVRLHDSI SEEGFHYLVF DLVTGGELFE  
 101 DIVAREYYSE ADASHCIQOI LEAVLHCHQM GVVHRDLKPE NLLLASKCKG  
 151 AAVKLADFGL AIEVQGDQQA WFGFAGTPGY LSPEVLRKEA YGKPVDIWAC  
 201 GVILYILLVG YPPFWDEDQH KLYQQIKAGA YDFPSPEWDT VTPEAKNLIN  
 251 QMLTINPAKR ITAHEALKHP WVCQRSTVAS MMHRQETVEC LKKFNARRKL  
 301 KGAILTTMLA TRNFSVGRQT TAPATMSTAA SGTTMGLVEQ AKSLLNKKAD  
 351 GVKPQTNSTK NSAAATSPKG TLPPAALEPQ TTVIHNPVDG IKESSDSANT  
 401 TIEDEDAKAR KQEIIKTTEQ LIEAVNNGDF EAYAFYFENL LAKNSKPIHT  
 451 TILNPHVHVI GEDAACIAYI RLTQYIDGQG RPRTSQSEET RVWHRRDGKW  
 501 QNVVHFCSCG PVAPLQ (SEQ ID NO:2)

# **FEATURES:**

## **Functional domains and key regions:**

[1] PDOC00001 PS00001 ASN\_GLYCOSYLATION  
 N-glycosylation site

Number of matches: 3

1 313-316 NFSV  
 2 357-360 NSTK  
 3 399-402 NTTI

[2] PDOC00004 PS00004 CAMP\_PHOSPHO\_SITE  
 CAMP- and cGMP-dependent protein kinase phosphorylation site

Number of matches: 2

1 48-51 KKLS  
 2 259-262 KRIT

[3] PDOC00005 PS00005 PKC\_PHOSPHO\_SITE  
 Protein kinase C phosphorylation site

Number of matches: 4

1 47-49 TTK  
 2 51-53 SAR  
 3 358-360 STK  
 4 367-369 SPK

[4] PDOC00006 PS00006 CK2\_PHOSPHO\_SITE  
 Casein kinase II phosphorylation site

Number of matches: 9

1 36-39 TGHE  
 2 51-54 SARD  
 3 79-82 SISE  
 4 94-97 TGGE  
 5 109-112 SEAD  
 6 262-265 TAHE  
 7 400-403 TTIE  
 8 401-404 TIED  
 9 485-488 SQSE

[5] PDOC00007 PS00007 TYR\_PHOSPHO\_SITE  
 Tyrosine kinase phosphorylation site

9-17 RFTDEYQLY

[6] PDOC00008 PS00008 MYRISTYL  
N-myristoylation site

Number of matches: 3

1 302-307 GAILTT  
2 332-337 GTTMGL  
3 390-395 GIKESS

[7] PDOC00100 PS00107 PROTEIN\_KINASE\_ATP  
Protein kinases ATP-binding region signature

20-43 IGKGAFSVVRRRCVKLCTGHEYAAK

[8] PDOC00100 PS00108 PROTEIN\_KINASE\_ST  
Serine/Threonine protein kinases active-site signature

132-144 VVHRDLKPENLLL

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	195	215	1.665	Certain
2	319	339	1.301	Certain

# BLAST Alignment to Top Hit:

```
>CRA|18000005245285 /altid=gi|5326757 /def=gb|AAD42035.1|AF078803_1
      (AF078803) calcium/calmodulin-dependent protein kinase II
      beta subunit; CAM2 [Homo sapiens] /org=Homo sapiens
      /taxon=9606 /dataset=nraa /length=542
      Length = 542
```

```
Score = 1047 bits (2678), Expect = 0.0
Identities = 516/542 (95%), Positives = 516/542 (95%), Gaps = 26/542 (4%)
Frame = +1
```

```
Query: 1      MATTVTCTRFTDEYQLYEDIGKGAFSVVRRCVKLCTGHEYAAKIINTKKLSARDHQKLER 180
      MATTVTCTRFTDEYQLYEDIGKGAFSVVRRCVKLCTGHEYAAKIINTKKLSARDHQKLER
Sbjct: 1      MATTVTCTRFTDEYQLYEDIGKGAFSVVRRCVKLCTGHEYAAKIINTKKLSARDHQKLER 60

Query: 181    EARICRLLKHSNIVRLHDSISEEGFHYLVFDLVTGGELFEDIVAREYYSEADASHCIQQI 360
      EARICRLLKHSNIVRLHDSISEEGFHYLVFDLVTGGELFEDIVAREYYSEADASHCIQQI
Sbjct: 61    EARICRLLKHSNIVRLHDSISEEGFHYLVFDLVTGGELFEDIVAREYYSEADASHCIQQI 120

Query: 361    LEAVLHCHQMGGVVRDLKPENLLLASKCKGAAVKLADFLAIEVQGDQQAQWFGAGTPGY 540
      LEAVLHCHQMGGVVRDLKPENLLLASKCKGAAVKLADFLAIEVQGDQQAQWFGAGTPGY
Sbjct: 121    LEAVLHCHQMGGVVRDLKPENLLLASKCKGAAVKLADFLAIEVQGDQQAQWFGAGTPGY 180

Query: 541    LSPEVLRKEAYGKPVDIWACGVILYILLVGYPFFWDEDQHKLYQQIKAGAYDFPSPWDWT 720
      LSPEVLRKEAYGKPVDIWACGVILYILLVGYPFFWDEDQHKLYQQIKAGAYDFPSPWDWT
Sbjct: 181    LSPEVLRKEAYGKPVDIWACGVILYILLVGYPFFWDEDQHKLYQQIKAGAYDFPSPWDWT 240

Query: 721    VTPEAKNLIQNMLTINPAKRITAHEALKHPWVCQRSTVASMMHRQETVECLKKFNARRKL 900
      VTPEAKNLIQNMLTINPAKRITAHEALKHPWVCQRSTVASMMHRQETVECLKKFNARRKL
Sbjct: 241    VTPEAKNLIQNMLTINPAKRITAHEALKHPWVCQRSTVASMMHRQETVECLKKFNARRKL 300

Query: 901    KGAILTTMLATRNFSVGRQTTAPATMSTAASGTTMGLVEQAKSLNKKADGVKQPQTNSTK 1080
      KGAILTTMLATRNFSVGRQTTAPATMSTAASGTTMGLVEQAKSLNKKADGVKQPQTNSTK
Sbjct: 301    KGAILTTMLATRNFSVGRQTTAPATMSTAASGTTMGLVEQAKSLNKKADGVKQPQTNSTK 360

Query: 1081   NSAAATSPKGTLPAALEPQTTVIHNPVDGIKESSDSANTTIEDEDAKARKQEIIKTTEQ 1260
      NSAAATSPKGTLPAALEPQTTVIHNPVDGIKESSDSANTTIEDEDAKARKQEIIKTTEQ
Sbjct: 361   NSAAATSPKGTLPAALEPQTTVIHNPVDGIKESSDSANTTIEDEDAKARKQEIIKTTEQ 420

Query: 1261   LIEAVNNGDFEAYA-----FYFENLLAKNSKPIHTTILN 1362
      LIEAVNNGDFEAYA                      FYFENLLAKNSKPIHTTILN
Sbjct: 421   LIEAVNNGDFEAYAKICDPGLTSFEPEALGNLVEGMDFHRFYFENLLAKNSKPIHTTILN 480

Query: 1363   PHVHVIGEDAACIAYIRLTQYIDGQGRPRTSQSEETRVWHRRDGKWQNVHFCSGAPVAP 1542
      PHVHVIGEDAACIAYIRLTQYIDGQGRPRTSQSEETRVWHRRDGKWQNVHFCSGAPVAP
Sbjct: 481   PHVHVIGEDAACIAYIRLTQYIDGQGRPRTSQSEETRVWHRRDGKWQNVHFCSGAPVAP 540

Query: 1543   LQ 1548
      LQ
Sbjct: 541   LQ 542 (SEQ ID NO:4)
```

# **Hmmer search results (Pfam):**

Model	Description	Score	E-value	N
PF00069	Eukaryotic protein kinase domain	306.2	3.9e-88	1
CE00022	CE00022 MAGUK_subfamily_d	293.8	1.3e-86	1
CE00359	E00359 bone_morphogenetic_protein_receptor	15.0	0.0015	1
CE00031	CE00031 VEGFR	0.9	2.1	1
CE00287	CE00287 PTK_Eph_orphan_receptor	-65.4	0.00046	1
CE00292	CE00292 PTK_membrane_span	-77.0	0.00018	1
CE00291	CE00291 PTK_fgf_receptor	-93.1	0.0021	1
CE00286	E00286 PTK_EGF_receptor	-132.2	0.0059	1
CE00290	CE00290 PTK_Trk_family	-161.3	0.00033	1
CE00016	CE00016 GSK_glycogen_synthase_kinase	-196.7	9.2e-06	1

## **Parsed for domains:**

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00359	1/1	132	186 ..	272	327 ..	15.0	0.0015
CE00031	1/1	133	205 ..	1068	1139 ..	0.9	2.1
CE00286	1/1	14	252 ..	1	263 []	-132.2	0.0059
CE00290	1/1	15	253 ..	1	282 []	-161.3	0.00033
CE00291	1/1	14	267 ..	1	285 []	-93.1	0.0021
CE00292	1/1	14	267 ..	1	288 []	-77.0	0.00018
CE00287	1/1	14	270 ..	1	260 []	-65.4	0.00046
PF00069	1/1	14	272 ..	1	278 []	306.2	3.9e-88
CE00022	1/1	10	305 ..	13	316 ..	293.8	1.3e-86
CE00016	1/1	1	343 [.	1	433 []	-196.7	9.2e-06

```

1 GAGCTGCTGT GTCTCTGTCC CCAGGGGCAG AGGGGCTGTG GGGTTGCAGG
51 CTCAGCGTCT GGGACTCTGG GGTGAAGGCT CAGCCATGCC CTGCAGACAC
101 CATGGGGCAG GGCTCAGACC TGTGCACCTG TCTCTTGCAA ACCACTGTTT
151 TCTCTGTTTT GTAACCCCCC ACCCAACCCC ACATAACACC TCTGGGTTTA
201 AACAACATGC ACCCTTGTGC CGGTCACTTC CCTGCAGCCG GAGAACCTGC
251 TTCTGGCCAG CAAGTGCAAA GGGGCTGCAG TGAAGCTGGC AGACTTCGGC
301 CTAGCTATCG AGGTGCAGGG GGACCAGCAG GCATGGTTTG GTGAGTGCCA
351 GGGGAGGGT GTGTTGGCTG GCAGTTGGCA GGGCAGGAGG TGATGCTGAC
401 AGCCCCTTGT GGCCTCTTCC CCTCTCTCTA GGTTTCGCTG GCACACCAGG
451 CTACCTGTCC CCTGAGGTCC TTCGCAAAGA GGCATATGGC AAGCCTGTGG
501 ACATCTGGGC ATGTGGTGAG GCCTGGCCTG AGTTGGTGCG GGGCAGGGCC
551 TCGGGTGTTC CAGGACTTCC CACCTACATC CTGGAGTGTG CAGTGGCCAG
601 CACGTCTTGC TCTCATCTGG GTTTATCTGT GTCAGACCTG CCCTTGAGCT
651 GCCCTGGCAG GGGTCTGCCC ACACAGCCAA GAGCCCCCTT TCCACCAGA
701 TTAGAATTGC TCACATGAAC CTGGCGCACC CCAGTGCTCG CCTGCGCTCA
751 GCAGAGGTCT GGTCCAGAAG TGTGGTGGGT GGATGGGAGT GGAGAAGAGA
801 GGTCAGGGGC TGTGGGCCA TGGGCAGGGC CACCTCCTTG GGTAGGGGTC
851 TCCTCCCACA GAGGTGGGGA GCAGCAGAGG GGCTTGACAT CACCCTCATC
901 CCTGTGATAG TGTGGGTGTG GGGCAGAGGT CAGGGGGCCG GCTGTGCCCT
951 TCTACCCAG TGTCTGCTGC ACAGGTGGGG GCAAAGGAAT GCTGAGGACC
1001 CCAATGCCCT CCCAGGGCCA CAGGAGCTAG GCAGTGAGGG TGCAGGGCAT
1051 GGGCTTATGC GACGTGGCA CCCTGCAAGT GGCTGCGGTG CTCACAGGCC
1101 CCATCCGCAG GGGTGATCCT GTACATCCTG CTCGTGGGCT ACCCACCCTT
1151 CTGGGACGAG GACCAGCACA AGCTGTACCA GCAGATCAAG GCTGGTGCCT
1201 ATGACGTGAG TGCACCAGCC CCTCTCTGAT GAGCTCCCTT CCTCCAGGTG
1251 TGGCCGGGTG AGGGCAGCGT GGGGAAGAGG TAGGAGTGGG GTGAAGCCAC
1301 CTGTGGCCAG GTCCTGGGTC CTGCTCTCCC AGATTCTGTG CTGGAGATGA
1351 AGCCCCCTTG AGAATTCTTG CCCCTGCCTG AGAGGGAGCT TCAGGCCCGG
1401 CCGGGGCGCT GTTTCCTTCT GCAGTTCCCG TCCCCTGAGT GGGACACCGT
1451 CACTCCTGAA GCCAAAAACC TCATCAACCA GATGCTGACC ATCAACCCTG
1501 CCAAGCGCAT CACAGCCCAT GAGGCCCTGA AGCACCCTG GGTCTGCGTG
1551 AGTCGCCCTT GGTGCCCATG GTGGGGAGGG GGCTCCTGTT GGAGATGGCC
1601 TCAGACCACT CCCCTGGCAA GGACCCCAAG AGGGTCTGTT TCCTGACATC
1651 CAAGAGCTCC CTGGGTGCTC CTGTGTGGCT CTGGCTTGGG
1701 ACATACCAGC ACGTTTGTGA GGCTGGGGC TTGGAAGGCA TTAGAGGGTA
1751 GAGGTGATCC CTCTCTCCCA ACTGCAGTCC TGTCTGTGAG GGGCAGAGTG
1801 GACGAGGCAA GGGAGAGACG AGTCTTGAAG TCCAGGCGG GTGGGGACAG
1851 ACAACCTTGG CCGCAATGGT GGCCGGTGGC TCTTGGCAAG TGGGGACCCC
1901 AGGGTGCCAC AAGCCTTGCC ACCCTGGCCT CTCCCCTGTG CCTCGGGCTC
1951 GGCTGCCATA TGACCACCCA TTTCCCCACA GCAACGCTCC ACGGTAGCAT
2001 CCATGATGCA CAGCAGGAG ACTGTGGAGT GTCTGAAAAA GTTCAATGCC
2051 AGGAGAAAGC TCAAGGTGAG GCCCTGGCCC CTAGTCCAG GCACGGCCAT
2101 GCTTCTCTGT GTCCCTCTGG GCTGGAGCAG GGGGGCCTTG GGGGTCTGG
2151 GCAGACCTAG GGGTTACTGC TGCCCCCAAG ACTGACTGTT AGCAAGTCCC
2201 AGACTGGATG CATCAGGTGA ACTCAGGCCA GCTTGGGAAT GAGTCCAGAG
2251 GGGCCCTGGG CCAGGTGTGG CTCCTCCTAG TTGTCTGTGC CACCTCCTAG
2301 CAGCCCTTGG AGGAGCTGTC CTGAAGCGCT CGCTGTGGGC TCCTACCCCG
2351 GGCTCTGCAG GCAGCACTCA CCCTCTGGCA GTCACACTGT TTAGTACAAG
2401 CAAGTCCGAA GCTTCCGGCT CAGACAGGTT TGGTAAGGAG AGCAGAGCCA
2451 CACACACTGG TCTTGGGTGG GCTGGGGGAG TTCTGGGAGG GAGGTGGGTC
2501 CCAGTAGGGT ATCCAACCTG CCTGCTTTGG TCAGGGCTGG CTCCGGTGAC
2551 CGCACACTGG CAGTCCCTCT ACTTGTGGGT TCCGGGATGG GGAATTGTTG
2601 CCTGACTGCC CTCTGCTGGT CTCTGAGCAG TTCTCCCCCG AAGCCCCAGG
2651 ACTGTTGCC TGTCTGAGCC TGTGAGGAAA AGAAGGGGCT GTCAGGGAGC
2701 TGGACCCAG AGGAGCTGCC GTGGTGACCA GCTGTTCTGG TGACCCCTGA
2751 GGCTTGAGGG GTCTTGAAGC AGCTAGAAGC TGTAAGTTGGT CAACAGGTTT
2801 AGGCCAGGG TGTGTGTAGT TCTGGAAATA GGTGATCTGT CTCAGTGCAG
2851 CTGCTGGCTT CCTGGAGCTC TTGCCTCTCT GGAAGGCTGA GGTCATGTCA
2901 GCCTCATGAC AATGAGGCTG AGCATCTGGG CAGGAGGACA GGGGTCTTAT
2951 CCTGGCCAGA AGCCAGCAGG GAACACTGAT GGGATAGCCC CGGTTTTATC
3001 TGTGTCTCTC CCCAGGGAGC CATCCTCACC ACCATGCTGG CCACACGGAA
3051 TTTCTCAGGT GAGCCTTTCT TCTCCAGGGA GACAGGCGCT GCCCCCTCCC
3101 TGCTGGCCCA CGCAGGAGAG CGCCTCCTTC CTCACCAGCC TCTCCACTCC

```

FIGURE 3, page 1 of 16

3151 TCCTCTGCGG CAGGCCTGCC CTCGGCGTCT GCCCTCAGCT CTGAGACCCA  
3201 CTGCCCCACCT GGGCCCCGCTG GGCTCCCACC TTGGGTGATA CCACAGGGTC  
3251 CAGCCCCCCG AGGCCATCAC CTTCTGTCTG GGTCTGTGTC CCTCCACCCC  
3301 CTGAACACGA GCGTCTGTGC TGCCCCACTG GGGCTCACAG CATCGTGTGT  
3351 GTCTGTCCAG GCGTTTGTGC GGCATCTATG TGGCCTCCTT GTCATTTTGA  
3401 GTGCTCTGAA CATTGTGTTT TGTGCGGGAG GTGGGCAGAA GGGATGCGGG  
3451 GTGATGCGGG AGGCTCGGGG GCCTCCTTCC AAGTTCTGGA TGAGCTGCAG  
3501 CCTCCTGTCC CGGCTGCTCA GGGTGGGTGG TTGGGAAGCA AGTTCTCTTG  
3551 GCAGGGGGGT GGGTCTGTG ATAGACCCCT GAGGCCAGG GCGCTGGCAG  
3601 ACCCATCGGG GCATGATGTT AGCCCCGAG TGGAGCCGGC AGCCAGGTCT  
3651 TGGACAAGCT GTACCTGTGG CTTCTCCGTC GTCCGACACT CCGTGTGCGA  
3701 GCGTCTGTGA TCCGTCTCTC TCGTTGTCCG TTTGCATCTG GTGCCCCCA  
3751 CCCGCCATCC TGTTACTTTT GCTGTGATGC TGTAATGCCG GGAACGCGTG  
3801 CACACGGTCA CACCAACACT AATAGGACTG TCCTGTCTGC TGTGTGCTCA  
3851 CCACACCCTT TGGGCATGAG AAGCCCCAC TGGGGTTTTT TAAGGAGAAA  
3901 GGAGGGCAAT GCTTTTCCGT GTCAATCAGT CCAATCTTGT TTCACTCTC  
3951 TTGAGCAAAG GATTCTGGAA CCATCTGTCA CCTAAACTTT AACTCTAATC  
4001 TTCTTCTGCT TCCTTTGTCT CTTTCTTCC CTTACCTCGC CCACCCCTCG  
4051 TCTGTGTCCG CCCACCCCTC CTTCCCTC GTCTCTAACC CGGTGCTAAC  
4101 AGTGGGAGA CAGACCACCG CTCCGGCCAC AATGTCCACC GCGCCTCCG  
4151 GCACCACCAT GGGGCTGGTG GAACAAGGTA GATGTGTCTC GACCAGCGTC  
4201 CCGCCCGCTC CCGCCCGTCC CTCCTGCCAG CATGCAGCCC CCTGTGTCAC  
4251 GCAGCCGCTG GCGGGGCTCC AGAGCCGCC CAGAGGCCGC CAGGCCCCCG  
4301 GGAGCCCTG CTCCCGTGTG GTCACATCCC AGCAGAGCCC ACCACAAGGG  
4351 CAGGGAGGCA GCCCCAAGG CTCCTCGCCT GTAAGAGGAG GGGCTGGGCT  
4401 AGGTGGCCCC TGGGCTACAC CAAGCCCTT TGGTCTTGGC CCCCAGGTCT  
4451 TGGGGTCCG GAGACCCCA TTAAGAATGG CCTGGGCCCC ACAGGGAGCC  
4501 ACTGGGCTG CTGCTGGGGG GTCTGAATCC TGAAAGGAGA GCCTTGAGGA  
4551 GCAGAGCAG AGAGGACAG GCCCTTGGG CAGACACACA CCCTGCCCTT  
4601 CTGGGGCCGC ATGGAGACGG TGGTCTGTGC TGCTGAGTCC TACACATGCA  
4651 TGTCTGCCCT GAGCATCCCC CCAGGACAAG CCGCTCTGGA GTGGGTGAGG  
4701 GTTTTATGCA CCCTGAGGAG ACTTTCAAG CTTCTCTTG GGTGTTTCT  
4751 GCAAAGTCTT CCTCCCTGG CCTCAAACCC TGTGAGGGAA AAGGCCGGCA  
4801 CTGGCCACCT GCTCCTCTGG GCTGTGCGG GCCAGAGCCC AGAGGCCCAA  
4851 GTTGGCTTCT GCCCACCTGC TGGCTTGTGA CCATGGGCAG ACCCATGAG  
4901 GGCTAGGCGA CCCCAGACC TCCTTGAGC TCCAGCCTGA GCTGAAGGCT  
4951 GGTGAGAGCT TAGGGCAGGC CAAGCTGACA ACGCCTGGCC ACAGAACACA  
5001 GAGGGCTACA GGGGTGACCC CAGATCCTCC CTGGGCTGAG CTGCTGAGTT  
5051 CCCTGTGCGT GCCTCCAACG TGGGCTGGGG ACCCGGCAGA GGTTCAGGG  
5101 TGCTGGAGAC TGCCTTCCC AGGCCTCCTC ATGACCCACA GGGTGAGCAG  
5151 CCTGGGCTTC CCAGCCAGG AACCTCCTT CTGGGGAGGC CCAGGGCGTC  
5201 CTCGGGGAGG GCAGTCTATT CTCCTCCCAT GAGCCAGTG GACGTGTCTA  
5251 GCAGGCAGCA CCCCAGGAGA GCCCTCCAC GTCTTCTCA TTTGACAGGC  
5301 CTTTCCAGAG CGCAGGCGGG AGGGGGCTGT GATTAGAAA GAGTGAGGCT  
5351 AGTGGCTTCT GGGGAGGCAC TGCTGCCAG GGGACAGTGC TGAGAGACAG  
5401 CTGCCTCTAC GCTGCCCTGT GCGCGGGCT CCCGCTGCAA TGCCCGCTG  
5451 TCTGCAAGTG AACGTGGGG GACGGTGCAT GAGGCCCTGC ATGTGTGGCT  
5501 CCACCCTGGG CGCCGAGAGC AGCTCTGTCC TGGAGGGTGG TCAGTGACATG  
5551 TGGACAGAGC CCAGCATGGC TGTCTGGGT GACCAGCTAA GGGGACAAGG  
5601 CAGAGGCAGG GCTGAGAGGA CCACCCATCC TGCTAGGTCA GCCCAGCTCA  
5651 GCCATATCAC ACGGCAGTGA GCATGGAGCT CAGTTCTCTG CCAATGGCAG  
5701 CTGAGTCTAG TACCATCCAG TCAGAGTCTG GTACCAGCCC ATGTGGCATA  
5751 GCCCCCTCG CCCGAGAGA GACCCGCTC GTCGAGTGTG CTTAGTTTG  
5801 GCCTCTGTGG TCTCTCTGC ATTGATCAGG TGTAAGGGCA TAGGAGACCC  
5851 AGTGTCCGGC CAGCTGCAGG GTGGCAGCAG TTGCCCCGGC CTGGAGACCC  
5901 GGGAATGGGC AGTGCCTTCC CAGGATGGAG GGCAGAGGGT CTCTCCTTGT  
5951 CCCACAGAGG CCTGCAGAAC CCCCACCCA GGTGTCTGAG ATGCCTGTGA  
6001 CTGCTCCGCC TACCCTGGGC TCCTGCGGCA CCTAACGCAT GCTTTGAACT  
6051 TGAGACACAG AAAGGAAGTT CCCGTGCCCT TGAATGCTAG TGTAGATGGG  
6101 CATCGACAGG ACTCTGGCCA CCGTGAATCT GGAGTTAGTC CCAGGCAGAG  
6151 ATGTGAAATG AGCAGCCCC CAAAAAATG TTGGCCGGGA GCCATGCACT  
6201 CAGGAGGGCC GGGCCCATGC ACCCCACACT GCGCCCAAG CGTGCACAAG  
6251 CGATTGTTTT AAAAGCGGGT TCACAAGGAA GGATGTTTGG GAACTGACTG

FIGURE 3, page 2 of 16



6301	AGACAACAGG	GACGTCTGCT	GCAGGGCTTC	CCAGAGCTCT	GATGGCAGCG
6351	TCGGCCTGAG	TCCTTCGAGG	AGGGCTGGTT	TGTACGTGGC	ATTGCTGCC
6401	CACCTGGACTG	TGAACCTCTG	TCTTTTATT	TCCCACTGCT	GCTGTGGTAC
6451	ATCTCCAGTA	GCATAGTTTG	GAAATGCAGG	TTTTGATAGA	CTCAAGGATC
6501	TAAATAGAAC	CCTCTTAGTA	CCAAGGACTG	TCCGGGGTCT	CTGCCAGCCC
6551	CGCCGATGGG	CCTAACTGTG	GTGCCTCCTT	TCCTGTGAGA	ATCTTCTGAG
6601	GACATGCCCG	GGGAAAGAGC	TCAGTTCTGC	TGCTGCCTAG	GGTGCCATGC
6651	TGGCCCCGGT	TCCAATGCAG	AGCCTAGCTG	GAAGTACCGC	TGGGTTGGCG
6701	GAGGCTACGT	GCCTGACTGT	CCCCTCGGGG	GTGGGGTGGA	ACTAGCCTTC
6751	TGAAACCGCC	TGCTTCAGTT	GGCCACAGCT	TTTTGAAATG	TGTGTTTCTG
6801	GAAGGGACTG	GGTCCCTTCC	TTGCCTGTTC	AGCTCCCCAC	GACAAATGTC
6851	CTCAAGGCGA	GGCTGGATGC	TTCTTTCCTC	AGGCTCCTAG	GAGGAGCCCG
6901	TCCCCCAGCT	GTGTCGGGCA	GCTGGTCACC	AGCAAGGACA	GGATCCCTCA
6951	GCTGCAGCCT	CAGGCTGGCT	GGCACTGGGC	GGGTGTTTCT	GGGATGAGTT
7001	GTGTGTACTG	GAGGTGGGAG	GGGAGCTGAG	AGGGTGGGAT	GCACAGACAG
7051	GAGAGGGGAC	TGTGGGGGTC	CTGGAACCCT	GAGTTCCAAG	TCTTCAGGAC
7101	TCTCCCTCCA	TAGCAAGTTA	CAGGGAAGCA	GATTTGAGCC	ACAGGGAAGC
7151	AGATTTGAGC	TGCAGCGAGG	GGGAGGGTTT	TCAGTCTGTG	CTATAGGGAA
7201	GTGGGCAGTC	GGCATTTCCTG	GTCCCTGGGAA	CTCACTGGGC	AGGGCTGCCT
7251	TGGGACATCA	GGGAGGTGCG	GCTGTGCTCA	GCTTCACCAG	GAGGGGCCTT
7301	AGGCCTGGGG	ACGGAGAGTG	ATGCCTGAGG	CCCCTCTACT	TCTCCATGGA
7351	TCCTGGGAGG	GACTCCTGGG	CTGGATACAA	AATTGTTGAG	AGTTAAGAGA
7401	TCTGTGAGGA	AGGGGAGGCT	GGGAATAGAA	AGTGTGTGCC	CACTGCACAT
7451	GGGGTCCGCA	GGGCCACGTG	CAGCCACTGC	GCAGGCACAA	CCCCAGTCCC
7501	CACAGAGCCC	AGGAGGGGCC	AGAGCCATGG	AGGAGGCAGC	ACTGGGCATT
7551	TGGACAGGGA	GGGGGTGGTC	AGCAGGCAGC	AGGCCCAGGC	CTGTCTATGC
7601	CCTGCGGGGT	GCAGCCTCCT	GATCTCCACG	GCAACCTGGA	GCACCCAGCG
7651	TCAGAACCAC	CGGGAGGGCT	TATGGAACAG	ATGTCCAGCC	CTGCAGAAGT
7701	TCTGGCTCAG	GAGGGCGGGG	TGGGCCTGGG	AATTGTCATT	TCTGACTGTA
7751	CAGGGCGATT	CTGCTGCTGC	TGCTGCTGCT	GGGGTTGGGG	GAGGATCCCA
7801	TTTGAGAAGC	GCTGCAGTCC	TAGGTTGAAA	CGTGCCTGTC	TGTCCCCACC
7851	CAGGCCTGCA	TGGGCAGCAC	GGGATCCCCA	GGCAGGAGGA	CCCAATTTC
7901	TGGCCTGGCC	AGCCAGGGTC	CTGGAGCCAG	GCGGTGGGGG	AGGGATGGGG
7951	GATTGCTGTG	CCACCTTCCT	TCCCGGCTTG	GCCCGGGGGC	AAGCATCCTC
8001	ACACTTCCCA	TGTCGTATC	CCCTTGGCTC	CAGCCTGGCT	GCCTCTCTAA
8051	CCCTGCTGTA	CCGGCTGGCC	GCATGGCCCT	GGCTCTTTT	GGTGAGCGTG
8101	GTCCAGGACT	GGTGACCTGT	GAGTCTGGG	CCCGCAGTCT	TGCGCCCCTG
8151	CCCGAACC	CACAAATCTT	GTTTTCTCTC	TCTCTCTTCC	TTCTCTACTC
8201	CCTCCCTTTC	TCACCTTTTC	TTTTCTGTAA	GGTAAGCTGA	CTTCTCTTTT
8251	TGGTTTTTTA	TTTATTTTAA	TTTTTTAGTT	CTGTAATTAA	AATCCTAACA
8301	GCCATTTAGG	GTGTGGGCAC	CGGGGGCTGG	GGCCAGGCC	CTCTGACCTC
8351	TGAGGGGGAA	TGCTGGGTGA	GGCAGGGGCC	CCGCTGCTGG	GACCAAGTAT
8401	CCTCAGGGGC	TTGTGGGCAG	AAAGGCCTGT	GCTGGCCCCA	GTCAAGTGCAC
8451	AGAAGCGGCC	CCAAGGCCAG	GGCTGCTGGG	CAGCTCGGAA	TGAGGGCGAG
8501	CAGGGCTGCC	CTTGGTGCCT	GAGCCAAGGA	GCCAATGGGA	CAGACCTCTG
8551	AGCCTGGGTG	CCAAGTATGA	GGTCTGAGAC	AGGGTGAGCG	CCTGGGCTGG
8601	GACAAGGCCC	TCTGAGTGGG	CGGCCAGCTG	CAGCCCACCC	ACCCCTACCC
8651	CAGGAAGGCA	GGGCCCGGGA	GGGCATGACC	TCTGGGGTGC	TGGCTCAGCT
8701	GCCCCACCC	CAACCTGACA	CCGCTAGTCC	TGAGTTCCCA	TCAGGGAGGA
8751	AGCAGCATCC	TGCCTTCCTC	TAGGAAGAGC	TTGCATGTGG	CCCAGAAGCC
8801	AAGGGGGCTC	CCCAGCACCC	ACGGGCATCT	CTGGGTCTGG	TCAGAGGAGA
8851	AATCTGGATG	CTTGACAGG	CCCCAGGGTC	ATGGAGGAGG	CTGGAGACAG
8901	GGCTGTCTTG	GGGTGATGGG	ATGGCCCCCC	CACCTGCTCA	GAGCCAGCCT
8951	GGGTGCTGGA	ACCACACTTG	CCTCAGGACC	CTGGGCTTGC	TCCTGGGGAA
9001	AGAGTGGGGT	CAGGCAAAGG	GGTGGGGTTG	CGCTGCAGCG	AGACCCAGGC
9051	CCATCACTCA	CCATACCTTC	TTCTTCCCCA	TGCAGCAGCC	AAGAGTTTAC
9101	TCAACAAGAA	AGCAGATGGA	GTCAAGGTGA	GGCTCCAGCC	GGGCCCTGTG
9151	GTGCCGGGGA	GCCAGAGGCC	TGCAGCTTCA	CCCCACGCC	CTGGGGCTCC
9201	TGCTCTGGAG	TCCCCCTCCC	CCCATGCCCT	GAGAGACACG	GGACAGGGAA
9251	TGGCGAGTGA	GGGGCTTCTC	CCACCTAAGA	GTTCCTCTTC	CCTCTCTCCA
9301	CAGCCCCAGA	CGAATAGCAC	CAAAAACAGT	GCAGCCGCCA	CCAGCCCCAA
9351	AGGGACGCTT	CCTCTGCCCG	CCCTGGTACT	GAGCTCCTCA	AATTCTGCCT
9401	CTCAGCCCC	CCTACGCCCC	TGGCTGTGTG	ATTGCCGCTG	GTCAGAGGGG

FIGURE 3, page 3 of 16

9451	GCCGGGTGAA	GGTGGGGTCT	GGCCCCGCCT	GGCTGTCTG	ACAGCACTCG
9501	CATGGCCCC	GCCCCTCATC	CCTCACCGGT	GGTGAAGTGG	AGAGAAGAGG
9551	CCACTGTTGT	GGGGGGCTCC	AATTCAGACA	GGTTTAGGAC	TGCTCTGGGG
9601	AGCCCCCTGGC	TGAGACCCAC	AGATGTTGGG	GTGCAGGGGA	GAGGCCCCAGC
9651	CTCCCACCCA	TGTGACTTGT	TGGATGTCTC	TCCAGGAGTG	TTCAGGAAGT
9701	CAGTGAGGCA	GAAATACCC	TCTCCCCACC	AGGACCCAC	CCTCAGCTCC
9751	TCCACCATCC	TCAAGCAGCC	GACCCACAGA	CCACTCCGAA	GGTCTGGCTT
9801	GGTGGGGCTG	GGCCAGGATC	TGCAGGGGGA	ACAGCCCAT	GTGGCACATT
9851	CCACGGCCCA	TGGGGAGACG	GGGCCACGGT	GGTGCAGTAG	AGAGGTGTCT
9901	AAGCCAGTGG	CAGCCAAGGG	GAGGGCTTGC	CGTCACCTCT	GTGTTCCCTC
9951	AGTGCTGTCT	TGTGGCTGCC	TGAGAGGCAG	GGCTTAGGGG	CTCCCTGCCG
10001	GGGAGGGGAG	GGGTCCCCAC	CATGCTCCGC	TCCAACGTGC	CCCCCTAGTG
10051	CCCCTTGCCC	TGGGGGCTCC	TACAGTGTAA	CCCTATAGCA	GTACTCCCAA
10101	GGATGTAAAG	TTGTGGCTGG	TGGGTGCCGG	CCTTCTGTCT	GGGGCGCTGT
10151	GCTGTGTCCC	CTCAGCTGTC	CTAAGAGCTT	TGGGGCTTGC	TGGCCCGTAG
10201	GTCCCCATAT	TTGCTGGAAG	CAGGCTTGGT	GTCCCTGAG	AACCCACAGC
10251	CAGGCTTCGG	GAGCCAGCCC	CAGACCCGCC	ACGGGAATAC	TGGGTTTGCC
10301	AAATGGCCAC	CTTGAGACCC	AGGAGAGGAG	AGCGGTCTCT	GGAGGGCCGA
10351	CTGTGCTCAGA	GCCAGCAGGC	CGTGGCTGGA	GGGTGGCCTG	GTGCAGCCTA
10401	CCTTAGGGCCT	TCCAGTGGCC	AGGGCAGCCC	ACGTGCCAGC	CTCACAGCCA
10451	GCCCCATCTC	GGACCCTGTC	CATCCCATGT	GCCACCGCCA	CCCCCATGAC
10501	ATCTTCAAAC	CTGTGCCCCC	CACCACGCTG	GGGCACAGGT	TCAGGCAGTA
10551	AAGGGTAGGG	AGAACCCCTC	AAGACCGAGC	CTGGCTTCTC	TGGCTCCCCAC
10601	ACACATTGTG	CAGCTTTGTCG	GGGCCCCACA	CGGTCCATCT	CCCACCCTGG
10651	ACACAGCAGC	CTCCGCCAGC	CTGGACAGAG	CCTCTGTCCA	TTCCATCCCT
10701	CCCGGCTGAC	CCAGGCTCCT	CCCCCAGCTG	CTCCACGCCG	CCTCCATCCC
10751	TGTCCCCCAC	TCTGCTCTGC	ACTTCTTTCT	CGCAGGCTCT	GGCCACCCAC
10801	ACCTCCTCTG	TCTCCCTGTT	CCCCTCCTGG	TGGTCTCCGC	TTCTCCTCT
10851	TCTCACTTTC	CCTCTCTTTC	CTTCCTCTGT	GTCTTCCTTC	TTCTGTAGGA
10901	GCCTCAAACC	ACCGTCATCC	ATAACCCAGT	GGACGGGATT	AAGGTACTGC
10951	CCACATTTCC	TCCTCCCGCT	TTCCCCAGCG	AGGAGGCTCC	AGGCCAGGAG
11001	AGAGGTCTGG	GCGAGCATTT	GTGCCAGAGT	GGAGGGCAGA	TGTCCCATGG
11051	CCCTGGCCGC	CCCTCCCCGC	AGTACGGTAG	GGCCCCAGTC	CGTCTTCGTG
11101	GGCAACAACA	GGACAGACTG	GCTCAGGCCC	CAGGCGCGCC	CCTGGAGGTG
11151	CTTGGCACAG	TTGCGCCCCG	TCCCCATGTG	GCCGACACTC	TCAGACCAGG
11201	GCTCTGCGTG	TCCCACCTAC	GGCAGGCGAT	AGGGCTTCTC	GAGGTCTGGA
11251	GCAGGGCCTG	CATCTCAGGA	GCTGCATCCT	TGGCCTCTCT	GGCTGTCTCT
11301	CACCCACCT	CCCTCACGTG	GCCCCCAGTG	CTTCTGTCTG	AGCAGACCCT
11351	CCCTCCTCTG	CTCCCCCTCT	TGCTCTGGCC	ATCAGCTCCC	ATCACATTGG
11401	CATCATCACT	CTGGGGCCAG	GGAAGGGGCT	GGCTCTCTGG	GGTGGTGGGA
11451	GGGATGGGGC	CAGCAGCCAA	GCCATTTCCA	GGACTTCCAA	AACAGCGCCA
11501	CTACACCCAA	CACGGCCCTC	CAGCCACAGT	CCCACCTAGG	CTCTGGCTCC
11551	TTACAGAGCC	CCCAGAGTGC	CTCTGTGGGG	ACCCCCACT	TCCTTCTGGC
11601	CAGTGCCACC	ACCCAGCCCA	TCATCAGAAG	ACATCTTTCT	CCATGGCAGG
11651	GACCAGGGGG	TCCAAGGGGC	ACCCATGGTG	CTAGGCACCA	GGGCCTGGGC
11701	ATTCTTCCCA	TCTGGCAGCT	GGGGATGGGT	GCCCCCTGGA	CCCGTGTGTG
11751	TCTGGGGTGG	GTCACTGTCT	CTGCAGGACT	CCTAAACAAC	CTTCTGGGCT
11801	GTGGTGAACT	CTGAGCCTGC	ACCTAAAAGA	CCTGTAGTTC	TGGTCTAGGG
11851	CCTCCAAGCA	GTGTCCAGGC	AGTGTCCAGA	CCAGGGGGCG	GTCCCCAGG
11901	GACCTTGTAA	GATGTTTCT	CTGAGGACGA	GAGCAGGCCT	CCTGGGGACC
11951	TGGGGGATGG	TCTTTTGAAG	GGCAGCAGCC	CTGGAGCAGG	GTGGGAGAGT
12001	CTGGGGCCAC	CTCTGCCCTC	TAAGGCCACC	TGAGAGGTGA	GGCCGGGGCC
12051	TGACTGGACG	TCCAGTCCCA	GAGGGGCAGG	TGCCCTGAGG	GAATGTGGGG
12101	GACAGGAATG	CTCTGCCTGG	GGCCAGGCCA	AGGTTCTCTG	AGCCCTGTGC
12151	GGATGCTGAC	AGCTCTCTGG	AACGCCCTCA	CCTGTATTTT	GGATGACACC
12201	GGCTGCTGCT	TCAATTGGAAC	CAGCCAGTCC	CATTGTGTTT	TACGTCTTGG
12251	AATTTCAAAA	AGCCCATTTT	CCTCTCTTGT	TAAAGAGTCA	GCTGAGCATA
12301	CCAGTCTCTC	TGCCAGGCTC	ATCTTGCTGG	GAGAAGTGGA	GCCCTCATGT
12351	GTTGGGGATG	CAGGTTGGCC	ACAGCACTAG	GGTGGCAGGG	CCGGCTCGG
12401	ACTCCGTGCC	AGCCTGTGCT	GGCTGCCGTG	AGAATGCCAC	CTGGTGAGGC
12451	GCGCCCTCCC	AGGGACCAAG	ACAGAACTGG	GTGTCTTCTC	CGGTCACTGG
12501	CGCATGAGGT	CCACAGAGCT	GGGGCCTGCG	AGCCGCCAGA	GGGCATGTCC
12551	CCTGAGCCCC	TGGCCTTTAA	GCCCCGTGGA	AGCAGCCGAG	GCAGAGATCA

FIGURE 3, page 4 of 16

GCTTCAGAGC	CTGGGCTGGT	CCTGACACAG	GCCCAGCCCT	GTCCACCTGC
CCTCAGCCAC	GTCCCACCTA	TCCTTGGCCG	CATCCTGACC	CGCTGCCCTCC
CGTGTTCCT	CAGGAGTCTT	CTGACAGTGC	CAATACCACC	ATAGAGGATG
AAGACGCTAA	AGGTACCTGC	ACTTGAGTCC	TTGCCCCCCC	AGCGGCCTTG
GCATTGCTGG	GTTGCTCTTT	GAGGTGGGTG	GGACTTGGGC	AGGGTCAACT
CTCCTGCGAC	GCCTAGTTTA	TGCATGTGTT	GAGGGGCTCA	GGGACCCTGT
AGCTGTAATC	CTGCTCCAAG	CCTGGGTGTC	AGGCCTGCCC	AGAGCGGAGA
AGCATGGCAG	AGATGACCGA	CAGCTGGGCA	GTCTCGGTCA	CCGCATCCAA
GTGAGGAAGC	CACGGCTTTG	CATGGAGGCA	GGTCTCCAC	ACCAGGACCC
TCACGGGGAA	ACAGGCCCAT	GGGTAGAATT	TGTTCCAAGA	TGCTGTCTCT
GTCTTAAAGC	TCCTTAAGCT	TGCGTTTCTG	TCCAGCATGC	ACTTGCCAAG
TGGCCGGGCA	GCTGGGTGAG	TGTTTCCGTG	TTTGCCCTTG	CTTAGCCAGG
AGTGTCTGTC	TGCGGTGGGT	TTCTGCACCA	CAGATTCCAG	GGCCCCCTCC
CTTGCTCACC	CAGGCCAATG	TCTTGTGTGT	TCCCCAAGAG	GCCCCCAGGG
CACCAGGCAC	TGGGGCATGC	TCCATGGATT	CTGCCGCCTC	CAGACCACCC
ACATGGGGCC	TCCTGACCCT	CATCGCTCAC	ACGGTCACCT	AATAAGCCTT
ATGCTGTTCT	CAGGGCTACC	CTGGTGCCCA	AAAAGGGTCA	GCCACTCTGC
CAGTTTAGGG	GAGAAAACTT	CTCACCTGTC	CAAAGCATAG	CCTTGCTCCT
GCCCCGCCTA	CCCAGCTATG	AACTGTCCC	TGAGCAGAGA	TGAGCACAGG
ACTTTGGGGC	CTGGATGCCG	GAGAGTGGGT	GTTTGTGTGA	TTCCCTTGCA
GTCTGGAACA	GGCCCCAAAG	GCAACAGCAT	GAAGGCTGTC	CAGAGGTTCT
CCATCACCTT	CAGCCGAGTG	GGGTGCTGAG	CAGTGAGGGA	GGGGACCTGG
GAGGGGGGCC	CAGCTGGAT	CCTGCAGGGG	AGAAGAGAAG	ACAGCCAGAA
GCCAGCAGCT	GTGGCTCAGA	TCTGAGCCCG	AGCAGCCTCT	CGAGGTGGAG
GCAGACACCC	CCCACCCAC	CCCGTGCA	AAGAAGCCTT	GCCAGCCTGC
CCTGAGGCTG	GTACAGAGTC	CAGGCAGGCT	CAGTGGCCAT	CATGCCCTTA
CGATGACTGT	CACTCCCTCT	CCGTGCGCCT	GGCCTCTGCT	GGCTCTGGCC
AGGGGTGGTG	ACAGCACTAG	GGTGGCAGGG	TGGCCTCTGA	CTCTGCGCCA
GCCTGCACTG	GCCTGTGCTG	CCCTGGCCTC	TGCTGGCTCT	GGCTCTGGCA
CCGGTCCCGT	GTTGGCTCCT	TCAGCCTTCA	CATACCTGCT	GCGGCCACCA
CAGGCCCAGG	ACCCCCACAG	GGTGCCACC	CCACCTCCAC	CCCAGGAGCC
CCAGGTATCC	AGCTGTCACC	CCCTCCCTCC	CTCCTGGCCT	CCCCCTGTCC
TTCTCCAGTT	GCCTTCTTTT	CTGCGGGCG	CACCACCCAC	CTGCCTGCCT
CACCTGTTTC	GCCTCAGCCC	CCAGGTCCC	CGACATCCTG	AGCTCAGTGA
GGAGGGGCTC	GGGAGCCCCA	GAAGCCGAGG	GGCCCTGCC	CTGCCCATCT
CCGGCTCCCT	TTAGCCCTCT	GCCAGCCCCA	TGTAAGTAGC	CTGGGTCTTG
CTGCTGTGGG	GGTCATGTTG	GAGGGCTGGC	AACCCCTTAG	AGGGGCCACT
CCAGAGCCGA	GGGCAGGCTG	AGCGTGGACC	CTGGCTCCAG	CCTCATCACC
CCACAATCCC	TCACTGGGGC	TTTCCAGGGT	GGCCCCAGCC	CATCGAGCCC
CACCTCTTTG	TGAGGAGGGC	CCTGGACCAC	TTTCTGTCTC	AAGGCCACTG
GGCAGGATGG	GAGGCCCTGG	AGGCTCGGGC	CTCAATTCCA	GTCTTCAGGG
TCGGTGCAGG	CCTCACTCCA	CCTCAGCTTG	CGGGCGGGGG	GGCTCCCTGC
TATTGAGGCA	GGCTCTGATT	CAGGGCCTGA	TCCCAGGGCC	CAAGGGGTCT
AGAACACGGG	ACCCCTCCCA	CTGGCCTCCT	CCGCCTTGCC	GCCGCCTCGT
GTGTCTGTCT	GCCTCATGTT	CACGTCTCAT	CTGTTCCACC	CCAGCCCCCA
GGGATCTCTG	ACATCCTGAA	CTCTGTGAGA	AGGGGTTTCA	GAACCCAGAA
AGCCGAGGGC	CCCCTCTCAG	CGGGGCCCCC	GCCCTGCCTG	TCTCCGGCTC
TCCTAGGCCC	CCTGTCTCTC	CCGTGTAAGT	AGTGGCCCCC	AGGCCTGCCG
CCTCTGTGTC	CGGACAGCTC	CCTGCGAATG	GCCGGCGCTC	AGCAGCTTCC
CACCTGCATG	CACGGCCCGA	CTACCCTGCC	CCGGCGCCGC	AGCCTGGAGT
CCTGCCCTGG	CGGGGCTTCC	TGTGGGCTCC	CATGCTAACC	AGCAGGGCAG
CTCCTGGCTT	CTCCCTAAGG	GGCCCAGACC	CCTCCACGGC	TCCTGCTCCC
ACTGCCACTC	CCCGCTCGCT	GTCCAGCCCC	AGGCCCTCT	CCAAAATGTC
TGTCCCAGCC	CTGGGCAGCC	CTGGCCCTC	CGAGGCCCCC	CATGCCCTTA
GGCCCTCTCT	GCTGATCACT	GTCCCAGCCC	CACAGACTTC	ACACCCACCC
AGGGGCCCTG	CCCATGGTGC	CCAGGAGCTG	CACTCAGGGC	CACCCTGGTT
CCTGATGTGG	CCCCAACCCC	TGAGCACCCCT	CCCTCAGTCT	AGGAGGCTGA
GGAAGGTGCC	AAAAGTGGAA	CCCCGACCAG	GGTCTCTGGA	GCTCACCAC
AAGGGGATAG	TACGGAGAAT	CATAAGCCTG	GCCTCTGCTG	ACCTGGGCTG
TCCTCATGGG	GCCAGGCCAG	GCCTCCTCTG	TAACGCCCGT	GACTCCCTCC
TCTCCCTGTA	ACCCCGTCCA	GCGTTCTCTA	AGGGCCACTT	ACCTGACAGC
TTCTTGCTGG	CCAGCAGCCT	CTCCCTGGAG	GGTGCCCTCT	GCCCCAGCA
GCTTCAGCCC	ACGCCACCCG	ACAGCCAGAG	CATCTGCCCT	TCACTCCTGC

FIGURE 3, page 5 of 16

15751 AGCCTCCTCT CCACGCACCA CGCTGTCCGC AGCAGCACCC TCTGTCCCCC  
15801 TGTCTCCCTC CGTCCCCCA TATCCCCCTC GGTCAGCCTA CAACCTCTCC  
15851 ACGTCCCCCT AAGTCCACGC TCTATCCCTA CATCCCCCTC TGTCCCCCAA  
15901 ATTCCCCCTT TTCCCTCATT TCCATTTTCC TCCCCAAACT CTGCTCTGCC  
15951 CCTCACATTC TCCCTCTGTC CCCCACACCC TCCTCTGTCC CCCACACCCT  
16001 CCTGTGTCCC CCACACCCTC CTCTGTCCCC CATATACCCC TCTGTCCCCC  
16051 ACACCACCTT TGGTCCCTTG CACGCCCTTT TCTGTCCCCC ACACCCTCTC  
16101 TGTTCCCTAC ACTCTCCCTC TGTCCTCCAG ACCCTCCTCT GTCCCCCACA  
16151 CTCCCTCTGT CCCCCACACC CCCTGTCCCC CACACTCTCC CTCTGCCCCC  
16201 CAGACCCTCC TCTGTCCCTT ACCTCCCTC TGTCCCCCAT ATCCCCCTCT  
16251 GTCCCCCACA CCCTCCTCTG TCCTCCACCC CCTGCCCCCC ATACCCTCTT  
16301 CTGTCCCCCA CACTTCCTCT GTCTTCCACA CCCCCTCTG TCCCCACAC  
16351 CCCCTCTGTC CCCCAGACTC TCCCTCTGTC CCCCACACTC CGTCTGTCCC  
16401 CCACACCTCC TGTCTTCCAC ACCCCCTTCT GTCCCCCACA CCCCCTCTGT  
16451 CCCCCATACT CTCCTCTGTC CCCCACCTCC CCTCTGTTCC CCACACCGCT  
16501 TCTGTCCCCC ACACCCTCTC TGTCTTCCAC TTCCCTCTG TCCCCACAT  
16551 CCCCTCTGTC CCCCACACCC CTCCTCTGTC CCCTGCACCC TCCTCTGTCC  
16601 CATGCACCTC TCTCTGTCCC CCACATCCCC CTCTGTCTCC CACACTCCCT  
16651 CTGTCCCCCA CATCCACCTT GGTCCCCCTA CGCACCCCCA TCCCCATGA  
16701 CCCCTTCTGT CCCCACACCC CCCTCTGTCT TCCACACCCC CCTCTGTCCC  
16751 CCACACCCAC CTTGGTCCCC TCATGCCCCC CATCCCTTAC ACCCCCACTT  
16801 TGTCCCCCCA CATGCCCTC TGTCCCCCAC GTTCCCTTCT GTCTCCCACG  
16851 TCTCTCCATC TTCCGTTTTC CCTCTCTGTC CCCCAGCTC CCCTCCATCC  
16901 CCCACATCCC CTTCTTTCCC CTATATCCCC TCTGTGCGCC CAGGTCCACC  
16951 ATCTTCCCCC CACACCCCCC CATTCTCCCT TCCTCCCTC TGTCCCCTTG  
17001 TGCCCCATCC CCCACATCTG CCTCTGTGCC CCTCAATCTC TGGCTTGGCT  
17051 GTCTGCCCAT GGTCTCTCTC CTGCGTGCCC CCGGTGCTG CCTTGTGTTC  
17101 ACGTCTCGTC TGTTCCGCCC CAGCCCCCAG GATCTCTGAC ATCCTGAACT  
17151 CTGTGAGGAG GGGCTCAGGG ACCCCAGAAG CCGAGGGCCC CTCGCCAGTG  
17201 GGGCCCCCGC CCTGCCCATC TCCGACTATC CCTGGCCCCC TGCCCCACCC  
17251 ATGTAAGTAG CACCTTGAGT GGCCGTGGCA GCGGCTGCCT GGAGGGGCTC  
17301 GGGGCGTGCG AGCCTGGCAG TGGTGCTCTG GGAAGGGCCA TTCTTGCGGA  
17351 GGAGGGCGGG GCACAGGATC CCTCTGCTGG GTCCCAGGGA ATTGCTTTGA  
17401 AGCACATGAA GGTGCCACTG GGTCTCAGAA AATGGAGGTT ATGGTTATGA  
17451 AGTGTGTATG ACATATGTGT ATAGGAAGAG CGTCCGAAAG AGCAGGTTTG  
17501 TTGCCGACCC CAGCATTCGC AACCCTGAGG TCCACAGCTT TCTCCTGATG  
17551 GGAGGGGAAT GGGTGGCAAA GGGTCTGCGC GTGTGGCAAG GGCTAGCACG  
17601 CCAGGAGCTG CTGGCTTGGG TCAAGGTGGA CCTGCTGGGC CGGGACAGAA  
17651 AAGTGTCACT CCCGGCTGTA GACGCTCTAG CATTAGAGCT GTCCAAGTCC  
17701 AGACAGCAGG GAGCAGGTGG GGATCGGGAG GCGCGGATCT GGGGGGCAGC  
17751 TGGGGCCAGG CTGAAACAGA GCGGGCGGGA CAGGAAGCAC AGGCTGGGCA  
17801 GCCTCCCCGG CCAGGGAGGA GCCAGGCTGG GCCACCTCCC GGTCTGTCTG  
17851 CCGACTACCC GCAGTATCAC TTACAGGGAT GGATGACATC CCAGGGCTGC  
17901 TGCCACCCCC ACCTGTGGGG AGACACCAGA CTGGGGGTGG TGTGGAGATA  
17951 CTCTTAGAGA AGAGGCTGCT GGGCCACGGG CTCGGCATGG CAGGGCAGTG  
18001 GCTAGGTAAG TACTTGAGGG ACAGGTGGGG TCTGCTTGCC ACCGTCCCCT  
18051 CTGCAGGCTG GGCCTGGGGG CTGCTGCAGG CGGCCAGGGC AGAAGGGTGT  
18101 GGGGAGAGTG AACCACACAG AGCAGCGGCT CGAGGAGGGG GATGCAGGCT  
18151 GCAGGCTCAA AGGGGCACTG GATCCACCCT GGGTGCCCCG GAGAGCAGGG  
18201 GGCAGCCCCC GGAGGGGTAC TCACCCCCAG AGCTTCTGTG GTCGGCTGAG  
18251 GACCCCCAGC AGGGGTTGAC TGAGGGGATC AGAGGCAAGC AGCTGAGGGG  
18301 AGAGGCCAGG TTCTTGATGC TGATAGGGTC GGGGTGCCTG GCGCACCAGA  
18351 ACTCAAGGAG GGAGGCATGG GGAGGGGCGG CCGTGCAGCT GGGGTGGGTG  
18401 CACCGCAGAG CCTCTGGGAG TGGTCAGAAC CCCCAGACCC TGCCACTTCT  
18451 ACAGCAGCTC ATCTGATTTT AAGGGGCTTG CTGCCCTTGC AGAAGTGGAG  
18501 GGGTGTGCCC AAAGGAGCCT GCCTGGAAGG TCACCCCATC AGGTGGGCAT  
18551 GACCCAGGCC CAGGACTGCA GCCTGCCCTC AAGGTCTGTG CAGTATCTGG  
18601 GGTGAGTCTT CTGAGGACAG GGCCAGGGT GGGTGTGGAG TGGCCAGCTC  
18651 GGGGTCGGGT GTCCAGGCTC ACCTTCAGGG GCCACAGCAC AGACCTGCCC  
18701 TTCCAGAGTC TTCCCTGAGC TTGGCTGGGG AGGAGGGGGC TGCAGGAAGG  
18751 AGCTGTGAGC AGGGCAGGAT GGAGATTCTG GTGGCCCTCC TGGGAGGGGC  
18801 TGGGCAGGGC TGGGAAAGGG GTGGGTGAGA TGTTCGGAA CTCAGGGAAA  
18851 GGAAGAGTCT GGGTACTGCC CTGGGGGCAC CTGGGCCAG GTGGCAGGTG

FIGURE 3, page 6 of 16



```

22051 GGGCAGGAGC TGCAGGCAGG CGCCCGGGAG TAGCCCGGGC ATGGGGGTGC
22101 GGGGCAACAG GGATGTCTGC AGGGGTAGCA TGTGGGCCCC GGAAGCAAGA
22151 CAGGTGGAGC CAGCCGGATG CGGCTCCTAT GAGAAAAGCG GGAACAAGA
22201 GACCACGCTC GTTCTTCCTG CTGCGGGGAC AGCCCTGGTC ATCGCTCCGG
22251 GGAACCTGCT AGCCTGCGCC GCACGTGGCC GCCCCCTGCT GCTTCCTCCT
22301 CCCC GGCTC CGGGTGGCCT TGCTGACGGC TCCTTCTCTG AGGCAGGTCT
22351 CTGCCTTCTC GCCTGGTGCC TGCACTCAGT AGCCCCCTCA CCAGAGCTGC
22401 TGGGTGAAGG AAGCACTAAG AACCAAGGC TCGGGAGGAG AGTGGGGCCG
22451 GGAAGCTGCA GGAAGCGCA GGGCCAGGCC TGGTGGGCCC AGGGGCTGGC
22501 TCACGGGAGG GCAGGAGGGA GACTGTGGCG GACAGCACGT GGGGCCAGGA
22551 GGTGACCTCC AAGTGGATTG TGGGTGGGTT TTTTGTCTC TTTCTGCATT
22601 TTCCAGGCAT TTTGTAATGT GGATAGAATA TTTCTGTTCT TCAAAAATAC
22651 TTTAGTTAAG AAAAATAAGA TGGAAGCTGT TGCACTTGAA AATGAGGAAG
22701 CCACTGGTGA TGCAGGGGGG GCGGCGGAGA GGACCTCTTC TGCAAATAGC
22751 GGCAGGAACA CGGCATGGAT GCAGCTCGCG CTCCCCCAGG CCCTCCCCTG
22801 GGCTGTGTGG AGGGGTCCGG GGGGAATGGG CCAGCGCCCA GTGGTCACCT
22851 GGCCATGTCT CCCACAGCC CGGAAGCAGG AGATCATTAA GACCACGGAG
22901 CAGTCTATCG AGGCCGTCAA CAACGGTGAC TTTGAGGCCT ACGCGTGAGT
22951 CCCTGGGGCT GGGGGGGGGC TGTGCAGGAC AAGGATGTGG GACCCTTGGG
23001 GGGGCTGCT CAGAGTCAGG GTTCCACGGG GCCCCCTCTC ACTTGATT
23051 GGCCCCCAGG AAAATCTGTG ACCCAGGGCT GACCTCGTTT GAGCCTGAAG
23101 CACTGGGCAA CCTGGTTGAA GGGATGGACT TCCACAGATT CTACTTCGAG
23151 AACCGTGAGT GAGGAAGCCC GGTGGGCAT GAGGGGGCGG TGCCCCCAGG
23201 AGAGCCTCTC GGCCCCCTCC AGGGACAGCA TGGTGGCTGC CTATGGAAGC
23251 CCTGTCCCCT CTGTGCCAG GGTGGCCAG CCACCTCTCC CCCGCCAGAG
23301 GCCATACCCA GCCCCAGAA TCCACTCTT GGAGGGGCC ATGCTGCTCC
23351 CAGGAGAGCC GAGCCTCCCC AATAAGGGGA GTTGAAGAG GGAAGGATT
23401 AGGCTGGTGG GGTGGAAGAC GGGCACCAGG GCAGTCATGG TAACCCGAGA
23451 CCCCCGCCCC CCCTGCTGTC CACAGTGCTG GCCAAGAACA GCAAGCCGAT
23501 CCACACGACC ATCCTGAACC CACACGTGCA CGTCATTGGA GAGGATGCCG
23551 CCTGCATCGC TTACATCCGG CTCACGAGT ACATTGACGG GCAGGGCCGG
23601 CCCC GCACCA GCCAGTCTGA GGAGACCCG GTGTGGCACC GCGCGACGG
23651 CAAGTGGCAG AACGTGCACT TCCACTGCTC GGGCGCGCCT GTGGCCCCGC
23701 TGCAGTGAAG GTGAGTGTTT TGTGCTAAGT GACAGCTGGG GCAGAGGGGT
23751 GGCGGTGGTG TGAGTGGCTG CAGCCTGGGG AGGCGATGGG GAGCGGTGGG
23801 GCCTGTGGCA GAGCCATGC CTGGGAAGTC CCTGAGCTTT CCTGGTGAGG
23851 CCACAGGAAT GATGTCAAAT TAGGGACCAC GGCAGGCTGG GTGTGGCAGG
23901 CCTCCCCAGA GGAAGTGGGA GCTGGTGAGG GCCTGAGCAG TCCACTCTGG
23951 CCAGAGCTGG GTGGGTGCA GGTGGATGGG CCCC GGCGAG CACAGTCTCTG
24001 GGCACCATGC CCTGTTTGTG AGGACTGTTA GAGCCCCAGA TGGGCGTTCC
24051 CCAGGTGGTG GGTGCAGCGG GCCCAGAGCC CAGTTTACAG GGGATAGTAG
24101 TAATTGGGTT GGGCACCTTG AACCTCTCTC CCGAGTGGGC CCTTTTCTGG
24151 ACTTTAACC TCTCTGCAGT GCGCATGGC AGACAGCAGA GCCTGGGGGT
24201 GGATGGGAGA GGGGGCTGCT GAGGAGCTGA CCCACCCGCC CCATTTCAGA
24251 GCTGCGCCCT GGTTCGCGG GACAGAGTTG GTGTTTGAG CCCGACTGCC
24301 CTCGGGCACA CGGCCTGCCT GTCGATGTT TGTGCTGCC TCGTTCCCTC
24351 CCCTGGTGCC TGTGCTGCA GAAAAACAAG ACCAGATGTG ATTTGTTAAA
24401 AAAAAAAAAA AAAAAACAAG ATGACGACGA CAACCACAAA
24451 AAAAAATTGAC ATCAGATGAA ATGAAAAAAA AAAAAACAA AAAAAACTAA
24501 AGGAAGGAAA AAGCTGTAAA AATCACTGGC ATTCGTGGGG CCACTCCCCA
24551 CCAAGCTCC ACGTGTGTCC GTCTGTGCTC CTGGCCTCTG GGGGACCAGC
24601 TGGGACATGA ACTTGTCTGC CAGGCCCCCG TCGCGTGCTG AACGGTGTTA
24651 GTTTGTAGGT AACGCACACA CCCCACACCT AAGGTGTCTG CATCCTCCTG
24701 CCAACGCATG GGCTCCACGT GGTGTGCTCG CTGGCTGTG TGAAGTCTAG
24751 CTGTCTCTTG GGAGGGGCTG TGGGGGCCCG CTGGGCTGCC TCCTTTCCCG
24801 CTAGTTGTGC CTGAGAGTTG CTGTTGTTCC TGCTTTCCCT TCCCTTCTCT
24851 TCATCCCCTG AAGGGCTAGG TGTGGGTTT CCGTGCCCG TATCCCCACA
24901 CACCCAGCAC GGACAACCT TCGGCAGAGC CCAGGCCGCG CCCTCACCCC
24951 CTGGAGTATT GAAACTGGAG TCCCGTCCCC AAGGCCTTCA GAGATGCCCC
25001 TACACACCCA GGGCTCCAGC TCTGGTCTCT CTGGGGGAGT AAAGTGCAAA
25051 GAGGGGCACA GCTTAGTTTT GGGCCTCTCG CCGAGCAAGA GACAGCACTG
25101 CTGGCTACAG CTCCAACACA GCCAGCTGTG GCAAGAGGAC TCTGCCTGGG
25151 CTGGCCCCC TCCTGTGTGA GGTGTCTGTC CCTTCTCTGC TGGCCAGCAG

```

FIGURE 3, page 8 of 16

25201 CAGATGCACT GGCAGCTCCC AACCTGTCTT CCGCCCCCTCG GCCCTCCCCC  
 25251 AGCCTGTTTC GCTTCTCTGC AGCCCGCAAG GGGGAGCAGA CTTTTGACAA  
 25301 AGGACTGCGG GCCTCGCTCA AGTCCCTGAG CCCCAGCTG AAGCTGGGAG  
 25351 GGGAGGCCAG GCTTTGTGTC TGGGCATATT CGTCTGCTGA TGGGGTTTGG  
 25401 GGAAGCCTGG GGCTTGGGGT TTGGTCGGGT GGTGCAGCTA GTGGCAGAGC  
 25451 GGGATCAGAG GTGGTGGCTG CCCAGCTTCT GGGCTGAGAC AAGGGTCTGT  
 25501 GCAGGGGTTT ACTGAAGTGG GAGTGCCTTT GGAATCTGGG CCGGGAGCAG  
 25551 AAGGGAGCAA AAGCTACAGT GGGAGCCAGC CTAGGGCACA TGGGAGGCGT  
 25601 GAGGGCAGTG CTGCCCCTGC AGTGTCAAGT GTGCCAGTGC CTGGCGGGC  
 25651 TGCAAGTGC GTGAGGGCAC CTCTAGGTG GGCCAGGGAT GCAGCTATGG  
 25701 AGATAAGCG GGCTGGGGAC AGAAACAGGT GGGCACAGGG CCCAGGACAC  
 25751 CAGCGGATGG AGGGCAGGGT CTAGCCCTGT GCTCCTGAGC GTCGGCTGCC  
 25801 TGGGTTTCGAG GCGGTGGGTC CCCGGCCCCC TGTGATGGTG TGTACCATGG  
 25851 GGGAGCTCGG GGACAGGGCA AGCCCGAGCA TGGTGGGGCT GCAGGGTGGG  
 25901 TCTGAAGCCA GGTGGGGTGG GGGTGGTCAC AAGCCCTGAC TGCAGAGGGT  
 25951 CAGGGGCTCC TGCCCCAGTG CCTGCCCCCT TCAATTAC ATGTTTTCA  
 26001 ACAAGGATTT TCTTTATCTT CCCCTACAAA TCAAGCCAAG GGAGGGGCAC  
 26051 AGAATGGGGA ACAGGACACA GGATCCTAAA CTCCAAGGGG ACTGTCCACC  
 26101 GATGAACACT CAGAGTGGAC ACCATCTTCC GTCCACGCTG TGCCAGGAC  
 26151 AGCTGTCCCC ATCCATGAAC ACAGGGTAAA CATCTGCCGG GCTCCGCACC  
 26201 AGTGCTCCC TGGGCCATGG GACAGCGGCA GGGCTACCA CGGACAGCAC  
 26251 GTGGCCAGC AGCCGGCCAC CCTGGCGTCC TGGGGCCTCC TCCCTCCTC  
 26301 TCCCTCTCAC CTTGTACCT CCACGGAGCT GCCTGTCTGG GATAATTTGG  
 26351 GGATTTTTTT TCTGGGGAT AATTCTTTTG CATGACCCCT AAAGAGCAAG  
 26401 CCACACCGGT CTGCTAGCTA GGTGTCCGCG GTGTGGTGGT GGCGCCGCT  
 26451 GGCCAGCGCT GCAAGGGGTC GGCTGCCAC GGTGCTGGCT GGCTCCCTC  
 26501 CCTCTCTCTT TTTGCTGAGT TTCATTGTCT TTTCTTTCTG AGCCTTGTA  
 26551 GTGTACAAAA ATTATTCTTA TTTTGTCTG TCTCGGAAA CTGCAAATAA  
 26601 AAGAAAAACA GGACAACTG CTTCAAGTGC AGCTGGGTGC TTTAGCTGGA  
 26651 ATCCTGCCGA CCTCTGCGC CAAAATACAG ACTCAAGCCC GGTCCCTGGC  
 26701 CAAGACCTTA CTTGGGCCCC TCCTCCAATG AAAGGTAGTG CTATGGGAGC  
 26751 CCTGAGCTGG CCCTGACAGT CCTGAGCCCC TCTAGGGTGA ACGGCTCACC  
 26801 CCAGGTAGGG CACTAGTCAT AGATCATAGC TCTACCAGCT GTCTCCACCT  
 26851 CTTCTCTGCG TCCTCTGAAG TCTTCTGGGC CCAGCGCTGT CCACCCTGAA  
 26901 TGCTGGAAC TAACTGGAT CCCAGCCCCC AACACCCCTG ACCTCTCCAT  
 26951 TCACCCCGG TGGCCGCTAA GGATGTGGCC AGGGCAGCCT CTGGGCAGGA  
 27001 AGGAGCCCCA GGACCAAGAC CTCTGGCTGT CTTGCTGTTT CTTCCGCCC  
 27051 CTGCTACATG TATTGGCTAT TCTGGATGCT GAGGACACAC AGTGACCACA  
 27101 GAGCCGGGCT CCACCCAGT GGATTATGCA GACAGATGGC ACGCAGGCCT  
 27151 GTGTGGACAT CAGCCTCGGG CACCAGACAT AGGCAAGGCG CAAGGTGATA  
 27201 CAGTAGGCAG CCACCATGGG GGCCAGGAG CTCCAGCAGA GGCCACACAA  
 27251 CCAGCCAGA ATCCAGGACA GAGAGCTGGA ATGGAGACAG GGAAGCCAGA  
 27301 TACCAGGCCA GACTGGCCAG GTGCTACAGG CCTGTGGGCC AGGCCAGGCT  
 27351 TGGGACTTTC GTCTGGGTG TGAAGGAGAC AGGCACCCCT GAGGCCTTCC  
 27401 CTCTGCATCT CCAGCCCAAG CTAAGCGCAA ACTCTTAGGT TGGAGTAAGG  
 27451 AGTAACCCCC TGCCAAGTTT CTCCTGTCTT CAGGCTCCAC CCACCACCTA  
 27501 TGCTGCCTGG CCCCATGGGG CACACGCTCA GGCCCAGCCT GGGAAAGCAA  
 27551 CTGCACCTGC CTGTGCTATG CTGGCCCTTC TCAGCCTCAA TGCCCTCCTC  
 27601 CCTCCCCGAC GCACCCCTCGT GGCCCCCGCT GGGCCCCCTG ATGCACCCCTC  
 27651 ATGTCTCCAT GGCAACCTGC TCAGAGTGTG GCCCTGCCCT TGGCTCCCCT  
 27701 CCACACCTGT GTCCCAGGCA GTGCCACGGC ACTTTCCTAA ACAGAAGGAT  
 27751 GGGCTTCAAA ACAGTCCCAG ACACTAAACA CACCTGCATT TTGGGTCCAA  
 27801 GTAACCTCTG ACAAGACGAG TGCCCCCTACA CACCCTCAGT CCTATCCACT  
 27851 ATGGGCAAGG AGCCTGAAG ATCCCCCAGA ACTGGCTAAA GCCCTCAGTC  
 27901 TCCTCCTCCA CCTGAGCAC CTTACGCGG CAGAGTGGCC CTGGATGTCA  
 27951 GCTTCTTGCT CCCCATGGTC TGCACCTGGA CAGGTGCTCT CAGGTGTGTG  
 28001 GGTGGGCAGG TGGCAGGTCC CAAGAGCCAG GTGCAAAGAA TCTAGGCCAG  
 28051 TGCCACAGAG TGCTGCAGTG TCTGTCCCCA GCATGGTATC TAGGGCTCCA  
 28101 CTTGCCTATC AGCTGTAATC GGAGGAGGCT TTCCAGGCCA GGCCTCCCCC  
 28151 AGGAAGGCTG CAGGCACTGC GGATCGTGC CCCTCACATG CATTATTCCT  
 28201 GAGGCCCTTC TGCAGATGCC ATCAGGGCAG CAACTCTGAT GAGGTATTAG  
 28251 GGCACAGCAC ACAGGGCTAA GCCACCCTGT ACTGGGCCAA GCGCTACAGG  
 28301 CAAAAAGGAC ACCACCGACG GGCATTTTCAT TCATCGCTTT TATTTTTATA

FIGURE 3, page 9 of 16



28351 TATTTTGTGAG AGGGAGCCTC ACTCTGTCGC CCAGGCTGGA GTGCAGTGGC  
 28401 GCGATCTTGG CTCACTGCAA CTTCTCCCTC CTGGGTTC (SEQ ID NO:3)

#### FEATURES:

Exon: 232-340  
 Intron: 341-431  
 Exon: 432-515  
 Intron: 516-1110  
 Exon: 1111-1205  
 Intron: 1206-1424  
 Exon: 1425-1547  
 Intron: 1548-1981  
 Exon: 1982-2065  
 Intron: 2066-3015  
 Exon: 3016-3058  
 Intron: 3059-4102  
 Exon: 4103-4177  
 Intron: 4178-9088  
 Exon: 9089-9126  
 Intron: 9127-9303  
 Exon: 9304-9375  
 Intron: 9376-10898  
 Exon: 10899-10943  
 Intron: 10944-12713  
 Exon: 12714-12762  
 Intron: 12763-17130  
 Exon: 17131-17133  
 Intron: 17134-22868  
 Exon: 22869-22944  
 Intron: 22945-23137  
 Exon: 23138-23154  
 Intron: 23155-23475  
 Exon: 23476-23705  
 Stop: 23706

#### CHROMOSOME MAP POSITION:

Chromosome 7

#### ALLELIC VARIANTS (SNPs):

DNA				Protein		
Position	Major	Minor	Domain	Position	Major	Minor
487	T	C	Exon	55	H	H
496	T	C	Exon	58	L	L
1662	T	C	Intron			
1785	T	A	Intron			
1889	A	T	Intron			
2416	C	T	Intron			
4698	A	G	Intron			
5424	C	T	Intron			
8722	C	A	Intron			
9982	G	A	Intron			
10951	C	T	Intron			
12603	T	C	Intron			
14583	C	T	Intron			
17290	T	C	Intron			
18188	C	T	Intron			
19911	A	G	Intron			
21328	C	A G	Intron			
21391	T	C	Intron			
22588	C	T	Intron			



22965	-	G	Intron			
23498	G	A	Exon	312	R	R
23663	T	C	Exon	367	S	S
25427	A	G	Beyond ORF (3')			
27727	C	T	Beyond ORF (3')			
27834	T	C	Beyond ORF (3')			
28336	G	A	Beyond ORF (3')			

Context:

DNA

Position

487

CACCTCTGGGTTTAAACAACATGCACCCTTGTGCCGGTCACCTCCCTGCAGCCGGAGAAC  
CTGCTTCTGGCCAGCAAGTGCAAAGGGGCTGCAGTGAAGCTGGCAGACTTCGGCCTAGCT  
ATCGAGGTGCAGGGGGACCAGCAGGCATGGTTTGGTGAGTGCCAGGGGCAGGGTGTGTG  
GCTGGCAGTTGGCAGGGCAGGAGGTGATGCTGACAGCCCCTTGTGGCCTCTTCCCCTCTC  
TCTAGGTTTCGCTGGCACACCAGGCTACCTGTCCCCTGAGGTCTTCGCAAAGAGGCGTA

[T, C]

GGCAAGCCTGTGGACATCTGGGCATGTGGTGAGGCCTGGCCTGAGTTGGTGCGGGGCAGG  
GCCTCGGGTGTTCAGGACTTCCCACCTACATCCTGGAGTGTGCAGTGGCCAGCACGTCT  
TGCTCTCATCTGGGTTTATCTGTGTGACACCTGCCCTTGAGCTGCCCTGGCAGGGGTCTG  
CCCACACAGCCAAGAGCCCCCTTCCACCCAGATTAGAATTGCTCACATGAACCTGGCGC  
ACCCAGTGCTCGCCTGCGCTCAGCAGAGGTCTGGTCCAGAAGTGTGGTGGGTGGATGGG

496

GTTTAAACAACATGCACCCTTGTGCCGGTCACCTCCCTGCAGCCGGAGAACCTGCTTCTG  
GCCAGCAAGTGCAAAGGGGCTGCAGTGAAGCTGGCAGACTTCGGCCTAGCTATCGAGGTG  
CAGGGGGACCAGCAGGCATGGTTTGGTGAGTGCCAGGGGCAGGGTGTGTGGCTGGCAGT  
TGGCAGGGCAGGAGGTGATGCTGACAGCCCCTTGTGGCCTCTCCCCTCTCTAGGTTT  
CGCTGGCACACCAGGCTACCTGTCCCCTGAGGTCTTCGCAAAGAGGCGTATGGCAAGCC

[T, C]

GTGGACATCTGGGCATGTGGTGAGGCCTGGCCTGAGTTGGTGCGGGGCAGGGCCTCGGGT  
GTTTCAGGACTTCCCACCTACATCCTGGAGTGTGCAGTGGCCAGCACGTCTTGCTCTCAT  
CTGGGTTTATCTGTGTGACACCTGCCCTTGAGCTGCCCTGGCAGGGGTCTGCCACACAG  
CCAAGAGCCCCCTTCCACCCAGATTAGAATTGCTCACATGAACCTGGCGCACCCAGTG  
CTCGCCTGCGCTCAGCAGAGGTCTGGTCCAGAAGTGTGGTGGGTGGATGGGAGTGGAGAA

1662

GAATTCCTGCCCCCTGCCTGAGAGGGAGCTTCAGGCCCGGCCGGGCGCTGTTTCCCTTCTG  
CAGTTCCTGCTCCCTGAGTGGGACACCGTCACTCCTGAAGCCAAAACCTCATCAACCAG  
ATGCTGACCATCAACCCTGCCAAGCGCATCACAGCCCATGAGGCCCTGAAGCACCCGTGG  
GTCTGCGTGAGTCGCCCTTGGTGCCCATGGTGGGGAGGGGGCTCCTGGTGGAGATGGCCT  
CAGACCACTCCCCTGGCAAGGACCCCAAGAGGGTCTGTTCTTGACATCCAAGAGCTCCC

[T, C]

TGGGTCCCCTGGGTGCTCCTTGTGGCCTCTGGCTTGGGACATACCAGCACGTTTGTGAGG  
CCTGGGGCTTGGAAAGGCATTAGAGGGTAGAGGTGATCCCTTCCCTCCCAACTGCAGTCTG  
TCTGTGAGGGGCAGAGTGGACGAGGCAAGGGAGAGACGAGTCTTGAAGTCCCAGGCGGGT  
GGGGACAGACAACCCCTTGCCGCAATGGTGGCCGGTGGCTCTTGGCAAGTGGGGACCCAG  
GGTGCCACAAGCCTTGCCACCTGGCCTCTCCCCTGTGCCTCGGGCTCGGCTGCCATATG

1785

CTGACCATCAACCCTGCCAAGCGCATCACAGCCCATGAGGCCCTGAAGCACCCGTGGGTG  
TGCGTGAGTCGCCCTTGGTGCCCATGGTGGGGAGGGGGCTCCTGGTGGAGATGGCCTCAG  
ACCACTCCCCTGGCAAGGACCCCAAGAGGGTCTGTTCTTGACATCCAAGAGCTCCCTTG  
GGTCCCCTGGGTGCTCCTTGTGGCCTCTGGCTTGGGACATACCAGCACGTTTGTGAGGCC  
TGGGGCTTGGAAAGGCATTAGAGGGTAGAGGTGATCCCTTCCCTCCCAACTGCAGTCTGTG

[T, A]

GTGAGGGGCAGAGTGGACGAGGCAAGGGAGAGACGAGTCTTGAAGTCCCAGGCGGGTGGG  
GACAGACAACCCCTTGCCGCAATGGTGGCCGGTGGCTCTTGGCAAGTGGGGACCCAGGGT  
GCCACAAGCCTTGCCACCTGGCCTCTCCCCTGTGCCTCGGGCTCGGCTGCCATATGACC  
ACCCATTTCCCAACAGCAACGCTCCACGCTAGCATCCATGATGCACAGACAGGAGACTGT  
GGAGTGTCTGAAAAAGTTCAATGCCAGGAGAAAGCTCAAGGTGAGGCCCTGGCCCCTAGT

1889

GTGGAGATGGCCTCAGACCACTCCCCTGGCAAGGACCCCAAGAGGGTCTGTTCTTGACA  
TCCAAGAGTCCCTTGGGTCCCCTGGGTGCTCCTTGTGGCCTCTGGCTGGGACATACCA  
GCACGTTTGTGAGGCCTGGGGCTTGGAAAGCATTAGAGGGTAGAGGTGATCCCTTCCCTC

FIGURE 3, page 11 of 16

CAACTGCAGTCCTGTCTGTGAGGGGCAGAGTGGACGAGGCAAGGGAGAGACGAGTCTTGA  
 AGTCCCAGGCGGGTGGGGACAGACAACCCCTTGCCGCAATGGTGGCCGGTGGCTCTTGGCA  
 [A, T]  
 GTGGGGACCCCAGGGTGCCACAAGCCTTGCCACCCTGGCCTCTCCCCCTGTGCCTCGGGCT  
 CGGCTGCCATATGACCACCCATTTCCCCACAGCAACGCTCCACGGTAGCATCCATGATGC  
 ACAGACAGGAGACTGTGGAGTGTCTGAAAAAGTTCAATGCCAGGAGAAAGCTCAAGGTGA  
 GGCCCTGGCCCCCTAGTCCCAGGCACGGCCATGCTTCTCTGTGTCCCTCTGGGCTGGAGCA  
 GGGGGCCTTGGGGGGTCTGGGCAGACCTAGGGGTACTGCTGCCCCAAGACTGACTGT

2416 TCTGGGCTGGAGCAGGGGGCCTTGGGGGGTCTGGGCAGACCTAGGGGTACTGCTGCCC  
 CCAAGACTGACTGTTAGCAAGTCCCAGACTGGATGCATCAGGTGAACTCAGGCCAGCTTG  
 GGAATGAGTCCAGAGGGGCCCTGGGCCAESTGTGGCTCCTCTAGTTGTCTGTGCCACCT  
 CCTAGCAGCCCTTGGAGGAGCTGTCTGAAGCGCTCGCTGTGGGCTCCTCACCCGGGCTC  
 TGCAGGCAGCACTCACCTCTGGCAGTCACACTGTTTAGTACAAGCAAGTCCGAAGCTTC  
 [C, T]  
 GGCTCAGACAGGTTTGGTAAGGAGAGCAGAGCCACACACACTGGTCTTGGGTGGGCTGGG  
 GGAGTTCTGGGAGGGAGGTGGGTCCCAGTAGGGTATCCAACCTGCCTGCTTTGGTCAGGG  
 CTGGCTCCGGTGACCGCACACTGGCAGTCCCTCTACTTGTGGGTTCGGGATGGGGACTT  
 GTTGCCTGACTGCCCTCTGCTGGTCTCTGAGCAGTTCTCCCCGAAGCCCCAGGACTGTT  
 GCCCTGTCTGAGCCTGTGAGGAAAAGAAGGGGCTGTGAGGGAGCTGGACCCAGAGGAGC

4698 GCTAGGTGGCCCCCTGGGCTACACCAAGCCCTTCTGGTCTTGCCCCCGAGGTCTGGGGGT  
 CCGGAGACCCCATTAAGAATGGCCTGGGCCCCACAGGGAGCCACTGGGCCTGCTGCTGG  
 GGGGTCTGAATCCTGAAAGGAGAGCCTTGAGGAGCAGAGCCAGAGAGGCAGAGGCCCTTG  
 GGGCAGACACACCCCTGCCCTCTGGGGCCGCATGGAGACGGTGGTCTGTGCTGCTGAG  
 TCCTACACATGCATGTCTGCCCTGAGCATCCCCCAGGACAAGCCGCTCTGGAGTGGGTG  
 [A, G]  
 GGGTTTTATGCACCCTGAGGAGACTTTCAAGGCTTCTCTTGGGTGTTTCTGCAAAGTC  
 CTCCTCCCCCTGGCCTCAAACCTGTGAGGGAAAAGGCCGGCACTGGCCACCTGCTCCTCT  
 GGGCTGTGCGGGCCAGAGCCAGAGGCCAAGTTGGCTTCTGCCCACCTGCTGGCTTGT  
 GACCAT

5424 CCTCCTCATGACCCACAGGGTGAGCAGCCTGGCCTTCCCAGCCAGAGAACCCTCCTTCTG  
 GGGAGGCCCAGGGCGTCTCGGGGAGGGCAGTCTATTCTCCTCCCATGAGCCAGTGGAC  
 GTGTCTAGCAGGCAGCACCCCGGAGAGCCCTCCCACGTCTTCTCCATTTGACAGGCCTT  
 TCCAGAGCGCAGGCGGGAGGGGCTGTGATTAGAAAAGAGTGAGGCTAGTGGCTTCTGGG  
 GAGGCACTGTGCCCAGGGGACAGTGCTGAGAGACAGCTGCCTCTACGCTGCCCTGTGCC  
 [C, T]  
 GGGGCTCCCGCTGCAATGCCCCGCTGTCTGCAAGTGAACGTGGGGCGACGGTGCAATGAGG  
 CCCTGCATGTGTGGCTCCACCTTGGGCGCCGAGAGCAGCTCTGTCTGGAGGGTGGTCAG  
 TGCATGTGGACAGAGCCAGCATGGCTGTCTGGGTGACCAGCTAAGGGGACAAGGCAGA  
 GGCAGGGCTGAGAGGACACCCATCCTGCTAGGTGAGGCTCAGCCAGCTCAGCCATATCACACGG  
 CAGTGAGCATGGAGCTCAGTTCTTGCCAATGGCAGCTGAGTCTAGTACCATCCAGTCAG

8722 AAGGCCTGTGCTGGCCCCAGTCAGTGACAGAAGCGGCCCCAAGGCCAGGGCTGCTGGGC  
 AGCTCGGAATGAGGGCGAGCAGGGCTGCCCTTGGTGCCTGAGCCAAGGAGCCAATGGGAC  
 AGACCTCTGAGCCTGGGTGCCAAGTATGAGGTCTGAGACAGGGTGAGCGCCTGGGCTGGG  
 ACAAGGCCCTCTGAGTGGGCGGCCAGCTGCAGCCACCCACCCCTACCCAGGAAGGCAG  
 GGCCCGGGAGGGCATGACCTCTGGGGTGTGGCTCAGCTGCCCCCACCACCTGACAC  
 [C, A]  
 GCTAGTCTGAGTTCCTCATCAGGGAGGAAGCAGCATCCTGCCTTCTCTAGGAAGAGCTT  
 GCATGTGGCCAGAAAGCCAAGGGGCTCCCCAGCACCCACGGGCATCTCTGGGTCTGGTC  
 AGAGGAGAAATCTGGATGCTTGCAGGAGCCCCAGGGTCATGGAGGAGGCTGGAGACAGGG  
 CTGTCTGGGGTGATGGGATGGCCCCCCCCACCTGCTCAGAGCCAGCCTGGGTGCTGGAAC  
 CACACTTGCCCTCAGGACCCTGGGCTTGTCTCTGGGAAAGAGTGGGGTCAGGCAAGGGG

9982 CCAGGAGTGTTTCAGGAAGTCAGTGAGGCAGAAGATACCCTCTCCCCACCAGGACCCACC  
 CTCAGCTCCTCCACCATCCTCAACAGGCCGACCCACAGACCACTCCGAAGGTCTGGCTTG  
 GTGGGGCTGGGCCAGGATCTGCAGGGGGAACAGCCCATAGTGGCACATTCCACGGCCCAT  
 GGGGAGACGGGGCCACGGTGGTGCAGTAGAGAGGTGTCTAAGCCAGTGGCAGCCAAAGGG  
 AGGGCTTGCCGTACCTCTGTGTTCCCTCAGTGCTGCTCTGTGGCTGCCTGAGAGGCAGG  
 [G, A]  
 CTTAGGGGCTCCCTGCCGGGAGGGGAGGGGTCCCCACCATGCTCCGCTCCAAGTGGCC

FIGURE 3, page 12 of 16

CCTCAGTGCCCCCTTGCCCTGGGGGCTCCTACAGGTGAACCCCTATAGCAGTACTCCCAAGG  
ATGTAAGAGTTGTGGCTGGTGGGTGCCGGCCTTCTGCTGGGGCGCTGTGCTGTGTCCCT  
CAGCTGTCTTAAGAGCTTGGGGCTTGTGGCCCGTAGGTCCCCATATTGCTGGAAGCA  
GGCTTGGTGTCCCCCTGAGAACCCAGGCCAGGCTTCGGGAGCCAGCCCCAGACCGCCAC

10951 ACAGCAGCACCTCCGCCAGCCTGGACAGAGCTCCTGTCCATTCCATCCCTGCCGGCTGAC  
CCAGGCTCCTCCCCAGCTGCTCCACGCCGCTCCATCCCTGTCCCCACTCTGCTCTGC  
ACTTCTTCTCGCAGGCTCTGGCCACCCACACCTCCTCTGTCTCCCTGTTCCTCTCTGG  
TGGTCTCCGCTTCTCTCTCTCACTTTCCTCTCTTCTCTCTCTCTGTCTCTCTCTC  
TTCTGTAGGAGCCTCAAACCACCGTCATCCATAACCCAGTGGACGGGATTAAGGTACTGC  
[C, T]  
CCACTTCTCTCTCCCGCTTCCCCAGGCAGGAGGCTCCAGGCCAGGAGAGAGGTCTGGG  
GCAGCATTTGTGCCAGAGTGAGAGGCAGATGTCCATGGCCCTGGCCGCCCTCCCCGCA  
GTACGGTAGGGCCCCAGTCCGTCTTCGTGGGCAACAACAGGACAGACTGGCTCAGGCC  
AGGCGCGCCCTGGAGGTGCTTGGCACAGTTGCGCCCGGTCCCCATGTGGCCGACACT  
CAGACCAGGGCTCTGCGTGTCCACCTACGGCAGGCAGTAGGGCTTCTGAGGTCTGGAG

12603 AGTCTCTCTGCCAGGCTCATCTTGCTGGGAGAAAGTGGAGCCCTCATGTGTTGGGGATGCA  
GGGTGGCCACAGCACTAGGGTGGCAGGGCCGGCCTCGGACTCCGTGCCAGCCTGTGCTGG  
CTGCCGTGAGAATGCACCCCTGGTGAGGGGCGCCCTCCCAGGGACAGCACAGAACTGGGT  
GTCTTCTCCGTCAGTGGCGCATGAGGTCCACAGAGCTGGGGCCCTGCAGCCGCCAGAGG  
GCATGTCCCTGAGCCCCTGCCCTTAAGCCCCGTGGAAGCAGCCGAGGCAGAGATCAGC  
[T, C]  
TCAGAGCCTGGGCTGGTCTGACACAGGCCAGCCCTGTCCACCTGCCCTCAGCCACGTC  
CCACCTATCCTTGGCCGCATCTGACCCGCTGCCTCCCGTGTTCCTCAGGAGTCTTCTG  
ACAGTGCCAAATACCACCATAGAGGATGAAGACGCTAAAGGTACCTGCATTGAGTCTTG  
CCCCCAGCGCCCTTGGCATTGCTGGGTGCTCTTTGAGGTGGGTGGGACTTGGGCAGG  
GTCAACTCTCTGCGACGCCAGTTTATGCATGTGTTGAGGGGCTCAGGGACCCTGTAGC

14583 ACATCCTGAGCTCAGTGAGGAGGGGCTCGGGAGCCCCAGAAGCCGAGGGGCCCCCTGCCCT  
GCCCATCTCCGGCTCCCTTTAGCCCCCTGCCAGCCCCATGTAAGTAGCCTGGGTCTGCT  
GCTGTGGGGGTGATGTTGGAGGGCTGGCAACCCCTAGAGGGGCCACTCCAGAGCCGAGG  
GCAGGCTGAGCGTGGACCCCTGGCTCCAGCCTCATACCCCAATCCCTCACTGGGGCTT  
TCCAGGGTGGCCCCAGCCCATCGAGCCCCACCTCTTGTGAGGAGGGCCCTGGACCACTT  
[C, T]  
CCTGCTCAAGGCCACTGGGCAGGATGGGAGGCCCTGGAGGCTCGGGCCTCAATTCCAGTC  
TTCAGGGTCCGTGCAGGCCTCACTCCACCTCAGCTTGCAGGGCGGGGGGCTCCCTGCTAT  
TGAGGCAGGCTCTGATTGAGGGCTGATCCCAGGGCCCAAGGGGTCTAGAACACGGGACC  
CCTCCCACTGGCCTCCTCCGCCTTGCCGCGCCCTCGTGTGTCTGTCTGCCTCATGTTAC  
GTCTCATCTGTCCACCCAGCCCCAGGGATCTTGACATCCTGAACTCTGTGAGAAGG

17290 CTGTCCCCTTGTGCCCCATCCCCACATCTGCCTCTGTGCCCCCTCAATCTCTGGCTTGGC  
TGCTGCCCCATGGTTTCTCTCCTGCGTGCCCCCGTGCCTGCCTTGTGTTACGTCTCGT  
CTGTTCCGCCCCAGCCCCAGGATCTCTGACATCCTGAACTCTGTGAGGAGGGGCTCAGG  
GACCCCAAGAGCCGAGGGCCCCCTGCCAGTGGGGCCCCCGCCCTGCCATCTCCGACTAT  
CCCTGGCCCCCTGCCACCCCATGTAAGTAGCACCTTGAGTGGCCGTGGCAGCGGCTGCC  
[T, C]  
GGAGGGGCTCGGGGCGTGCGAGCCTGGCAGTGGTCTCTGGGAAGGGCCATTCTTGCGGA  
GGAGGGCGGGGCACAGGATCCCTCTGCTGGGTCCCAGGGAATTGCTTTGAAGCACATGAA  
GGTGCCACTGGGTCTCAGAAAATGGAGGTTATGGTTATGAAGTGTGTATGACATATGTGT  
ATAGGAAGAGCGTCCGAAAGAGCAGGTTTGTGCGGACCCAGCATTCGCAACCCCTGAGG  
TCCACAGCTTTCTCTGATGGGAGGGGAATGGGTGGCAAAGGTCTGCGCGTGTGGCAAG

18188 ATCCCAGGGCTGCTGCCACCCACCTGTGGGGAGACACCAGACTGGGGGTGGTGTGGAG  
ATACTCTTAGAGAAGAGGCTGCTGGGCCACGGGCTCGGCATGGCAGGGCAGTGGCTAGGT  
AAGTACTTGAGGGACAGGTGGGGTCTGCTTGCCACCGTCCCCCTCTGCAGGCTGGGCTGG  
GGGCTGCTGCAGGCGGCCAGGGCAGAAAGGTGTGGGGAGAGTGAACCCACAGGAGCAGCG  
GCTCGAGGAGGGGATGCAGGCTGCAGGCTCAAAGGGGCACTGGATCCACCCCTGGGTGCC  
[C, T]  
GAGAGAGCAGGGGGCAGCCCCCTGGAGGGGTACTACCCCCAGAGCTTCTGTGGTCCGGCTG  
AGGACCCCCAGCAGGGGTTGACTGAGGGGATCAGAGGCAAGCAGCTGAGGGGAGAGGCCA  
GGTTCTTGATGCTGATAGGGTCCGGGTGCTGGGCGACCAGAACTCAAGGAGGGAGGCAT  
GGGGAGGGGCCCGCTGCAGCTGGGGTGGGTGCACCGCAGAGCCTCTGGGAGTGGTCAGA

FIGURE 3, page 13 of 16

ACCCCCGACACCTGCCACTTCTACAGCAGCTCATCTGATTTTAAGGGGCTTGCTGCCCTT

19911 AGCACGGTTACCACTCTTGATTGGAACCTGACCATGCATCTCCTCTTCTGTTTACTTCA  
CGCTTTCTCTTCCCATCAACTCCCATTTTAATTACAATTTGTTTAAAAGCACTGCATATT  
ACTTCATTAAACAGAAGATTAGTTTCACTTACCATTAGTGTAAAGTGACTATAGAACCAA  
AGCAGACTGGAAACCAAATGACATAATGTCATTCTCTCTCCATTCCAGCTGCCTGCTGC  
TGTGCGCCTGAGAACCCTGTGGAGTGGGAGGGGACAGTGTCTCTGTACATTAGAAAGGG  
[A, G]  
GGTTAACTAAGTGACAGGAGGTGTTTGGGACATGTGGACACCAGACTTCTCTCTTGATGC  
AAGGAGGGCAGAGCCAGGCAGCCTAGTGGGGGCTGGCTTGGGGGCTGCTGGAAGGACTGG  
CTACAGGTGGAAGAGAGGTGACACCTGAAGCTTGGGGCCACCTCCAGGAAAGGACAGGTG  
AAAGTGGAGGCATGAGGCAGGGGACAGGCAGGTGCCAGGCAGAGGTGGAGAGGAGGCAG  
GAACATAGCAGCTGGGGCGGGGCGGGCCCTCAAGTGTATATGCTACTTCTCTGGGGCC

21328 GCTGGGCACAGTGGCTCATACCTGTAATCCCAGCACTTGGGAGGCCGAGGTGGGCAGAT  
CACTTGAGGTTAGGAGTTTGAGACCAGCCTGGCCAATATGGTGAAACCTCATCTCCACTA  
AAAATATACACACAAAAAATTAGCTGGGTGTGGTGGTGTGCACCTGTAGTTCCAGCTAC  
TCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGTCAGAGACTGCAGTGAGCCGA  
GATCATGTCACTGCCTCCAGCCCCGGTGACAGAGTGAGACTCCATCTAAAAA  
[C, A, G]  
AATTCCTCTCTGGGAATTTAGACCACAGACAGGTGTCATGTATGTGGCCGTGGAGGC  
AGCACTCACAGCAAAGAGTGGAAACGTCAACACAGGGCCTGCCTTCTGGTGAAAATGGTG  
TCCTGCAGGGCGGGCAGCTGTTTGGGGCAGGTGTCCAGGTGCGGCCCTGCAGCAGCCTG  
AGGGTCACAGAGCGCAGTGTGGGAGTGCAGAGACTTCCCCACAGGGAGAGTTCCAGG  
AACCTGCTTCCGGTGCACTTCTGGGGGTTGAGTTTTTCCACGGACGAATTACTTTGAG

21391 TTGAGGTTAGGAGTTTGAGACCAGCCTGGCCAATATGGTGAAACCTCATCTCCACTAAAA  
ATATACACACAAAAAATTAGCTGGGTGTGGTGGTGTGCACCTGTAGTTCCAGCTACTCG  
GGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGTCAGAGACTGCAGTGAGCCGAGAT  
CATGTCACTGCCTCCAGCCCCGGTGACAGAGTGCAGACTCCATCTAAAAA  
TTCCCTCTCTGGGAATTTAGACCACAGACAGGTGTCATGTATGTGGCCGTGGAGGCAG  
[T, C]  
ACTCACAGCAAAGAGTGGAAACGTCAACACAGGGCCTGCCTTCTGGTGAAAATGGTGTCC  
TGCAGGGCGGGCAGCTGTTTGGGGCAGGTGTCCAGGTGCGGCCCTGCAGCAGCCTGAGG  
GTCACAGAGCGCAGTGTGGGAGTGCAGAGACTTCCCCACAGGGAGAGTTCCAGGAAC  
CTGCTTCCGGTGCACTTCTGGGGGTTGAGTTTTTCCACGGACGAATTACTTTGAGAAA  
CCACTGTTACTCGTGTGTATAGGTGAGCGTGCCTGTGCATGTGTGTTCTGTGTGTGAGTG

22588 GCTGCTTCCTCTCCCCGGCCTCCGGGTGGCCTTGCTGACGGCTCCTTCTCTGAGGCAGG  
TCTCTGCCTTCTCGCCTGGTGCCTGCACTCAGTAGCCCCCTACCAGAGCTGCTGGGTGA  
AGGAAGCACTAAGAACCCAGGCTCGGGAGGAGAGTGGGGCCGGAAGCTGCAGGGAAGC  
GCAGGGCCAGGCCTGGTGGGCCAGGGGCTGGCTACGGGAGGGCAGGAGGGAGACTGTG  
GCGGACAGCACGTGGGGCCAGGAGGTGACCTCCAAGTGGATTGTGGGTGGGTTTTTTGTC  
[C, T]  
TCTTCTGCATTTTCCAGGCATTTGTAATGTGGATAGAATATTTCTGTTCTTCAAAAAT  
ACTTTAGTTAAGAAAAATAAGATGGAAGCTGTTGCACTTGAAAATGAGGAAGCCACTGGT  
GATGCAGGGGGGGCGGGCAGAGGACCTCTTCTGCAATAGCGGCAGGAACACGGCATGG  
ATGCAGCTCGCGTCCCCCAGGCCCTCCCCCTGGGCTGTGTGGAGGGGTCCGGGGGAATG  
GGCCAGCGCCAGTGGTCACTGGCCATGTCTCCCCACAGCCCGGAAGCAGGAGATCATT

22965 ATAAGATGGAAGCTGTTGCACTTGAAAATGAGGAAGCCACTGGTGATGCAGGGGGGGCGG  
CGGAGAGGACCTCTCTGCAAATAGCGGCAGGAACACGGCATGGATGCAGCTCGCGCTCC  
CCCAGGCCCTCCCCCTGGGCTGTGTGGAGGGGTCCGGGGGAATGGGCCAGCGCCAGTGG  
TCACCTGGCCATGTCTCCCCACAGCCCGGAAGCAGGAGATCATTAAGACCACGGAGCAGC  
TCATCGAGCCCGTCAACAACGGTGACTTTGAGGCCTACGCGTGAAGTCCCTGGGGCTGGGG  
[-, G]  
GGGGCTGTGCAGGACAAGGATGTGGGACCTTGGGGGGGCTGCTCAGAGTCAGGGGTCC  
ACGGGGCCCCCTCCTCACTTGGATTTGGCCCCCAGGAAAATCTGTGACCCAGGGGTGACCT  
CGTTTGAGCCTGAAGCACTGGGCAACCTGGTTGAAGGGATGGACTTCCACAGATTCTACT  
TCGAGAACCGTGAGTGAGGAAGCCCGGTGGGCATGAGGGGGCGTGCCCCCAGGAGAGC  
CTCTCGGCCCTCCAGGGACAGCATGGTGGCTGCCTATGGAAGCCCTGTCCCCTCTGTG

23498 CCCGCCAGAGGCCATACCCAGCCCCCAGAATCCCCTCTTGAGGGGCCCATGCTGCTCC

FIGURE 3, page 14 of 16

CAGGAGAGCCGAGCCTCCCAATAAGGGGAGTTGAGAGAGGGAAAGGATTAGGCTGGTGG  
GGTGAAGACGGGCACCAGGGCAGTCATGGTAACCCGAGACCCCCCGCCCTGCTGTC  
CACAGTGCTGGCCAAGAACAGCAAGCC  
[G, A]  
ATCCACACGACCATCCTGAACCCACACGTGCACGTGATTGGAGAGGATGCCGCTGCATC  
GCTTACATCCGGCTCACGCAGTACATTGACGGGCAGGGCCGGCCCGCACCAGCCAGTCT  
GAGGAGACCCGCGTGTGGCACC CGCGACGGCAAGTGGCAGAACGTGCACTTCCACTGC  
TCGGGCGCGCCTGTGGCCCCGCTGCAG

23663

GCCTCCCAATAAGGGGAGTTGAGAGAGGGAAAGGATTAGGCTGGTGGGTGGAAGACGG  
GCACCAGGGCAGTCATGGTAACCCGAGACCCCCCGCCCTGCTGTCCACAGTGTGGC  
CAAGAACAGCAAGCCGATCCACAGACCATCCTGAACCCACACGTGCACGTATTGGAGA  
GGATGCCGCTGCATCGCTTACATCCGGCTCACGCAGTACATTGACGGGCAGGGCCGGCC  
CCGCACCAGCCAGTCTGAGGAGACCCGCGTGTGGCACC CGCGACGGCAAGTGGCAGAA  
[T, C]  
GTGCACTTCCACTGCTCGGGCGCGCCTGTGGCCCCGCTGCAGTGAAGGTGAGTGTCTGT  
GCTAAGTGACAGCTGGGGCAGAGGGGTGGCGGTGGTGTGAGTGGCTGCAGCCTGGGGAGG  
CGATGGGGAGCGGTGGGGCCTGTGGCAGAGCCCATGCCTGGGAAGTCCCTGAGCTTTCCT  
GGTGAGGCCACAGGAATGATGTCAAATTAGGGACCACGGCAGGCTGGGTGTGGCAGGCCCT  
CCCCAGAGGACTGGGGAGCTGGTGAGGGCCTGAGCAGTCCACACTGGCCAGAGCTGGGTG

25427

TGTGGCAAGAGGACTCTGCCTGGGCTGGCCCCCTCCTGTGTGAGGTGTCTGTCCCTTCT  
CTGCTGGCCAGCAGCAGATGCACTGGCAGCTCCCAACCCTGTTTCCGCCCCCTCGGCCCTC  
CCCCAGCCTGTTTCGGCTTCTCTGCAGCCCCGCAAGGGGAGCAGACTTTTGACAAAGGACT  
GCGGGCCTCGCTCAAGTCCCTGAGCCCCAGCTGAAGCTGGGAGGGGAGGCCAGGCTTTG  
TGTCTGGGCATATTCTGCTGCTGATGGGGTTTGGGAAGCCTGGGGCTTGGGGTTTGGTC  
[A, G]  
GGTGGTGACAGCTAGTGGCAGAGCGGGATCAGAGGTGGTGGCTGCCAGCTTCTGGGCTGA  
GACAAGGGTCTGTGCAGGGGTTTACTGAAGTGGGAGTGCCTTTGGAATCTGGGCCGGGAG  
CAGAAGGGAGCAAAGCTACAGTGGGAGCCAGCCTAGGGCACATGGGAGGCGTGAGGGCA  
GTGCTGCCCGTGCAAGTGTGAGGTGTGCCAGTGCCTTGGCGGGCTGCAGTGCCTGTGAGGG  
CACCTTCTAGGTGGGCCAGGGATGCAGCTATGGAGATAAGGCGGGCTGGGGACAGAAACA

27727

GCAAACCTCTTAGGTTGGAGTAAGGAGTAACCCCTGCCAAGTTTCTCCTGTCTCAGGCT  
CCACCCACCACCTATGCTGCCTGGCCCCATGGGGCACACGCTCAGGCCCAGCCTGGGAAA  
GCAACTGCACCTGCCTGTGCTATGCTGGCCCTTCTCAGCCTCAATGCCCTCCTCCCTCCC  
CGACGCACCCCTCGTGGCCCCCGCTGGGCCCCCTGATGCACCCCTCATGTCTCCATGGCAAC  
CTGCTCAGAGTGTGGCCCTGCCCTTGGCTCCCCTCCACACCTGTGTCCAGGCAGTGCCA  
[C, T]  
GGCACTTTCCTAAACAGAAGGATGGGCTTCAAACAGTCCCAGACACTAAACACACCTGC  
ATTTTGGGTCCAAGTAACCTTCTGACAAGACGAGTGCCCTACACACCCTCAGTCTCTATCC  
ACTATGGGCAAGGAGCCTGAAGGATCCCCAGAACTGGCTAAAGCCCTCAGTCTCCTCCT  
CCACCCTGAGCACCTTACGCGGCAGAGTGGCCCTGGATGTGAGCTTCTTGCTCCCCATG  
GTCTGCACCTGGACAGGTGCTCTCAGGTGTGTGGGTGGGCAGGTGGCAGGTCCCAAGAGC

27834

CCAGCCTGGGAAAGCAACTGCACCTGCCTGTGCTATGCTGGCCCTTCTCAGCCTCAATGC  
CCTCCTCCCTCCCCGACGCACCCCTCGTGGCCCCCGCTGGGCCCCCTGATGCACCCCTCATG  
TCTCCATGGCAACCTGCTCAGAGTGTGGCCCTGCCCTTGGCTCCCCTCCACACCTGTGTC  
CCAGGCAGTGCCACGGCACTTTCTAAACAGAAGGATGGGCTTCAAAACAGTCCCAGACA  
CTAAACACACCTGCATTTTGGGTCCAAGTAACCTTCTGACAAGACGAGTGCCCTACACAC  
[T, C]  
CTCAGTCTCTATCCACTATGGGCAAGGAGCCTGAAGGATCCCCAGAACTGGCTAAAGCCC  
TCAGTCTCCTCCTCCACCCTGAGCACCTTACGCGGCAGAGTGGCCCTGGATGTGAGCTT  
CTTGCTCCCCATGGTCTGCACCTGGACAGGTGCTCTCAGGTGTGTGGGTGGGCAGGTGGC  
AGGTCCCAAGAGCCAGGTGCAAGAATCTAGGCCAGTGGCCACGAGTGTGTCAGTGTCTG  
TCCCCAGCATGGTATCTAGGGCTCCACTTGCCTATCAGCTGTAATCGGAGGAGGCTTTC

AGAATCTAGGCCAGTGCCACGAGTGCTGCAGTGTCTGTCCCCAGCATGGTATCTAGGG  
CTCCACTTGCCATATCAGCTGTAATCGGAGGAGGCTTCCAGGCCAGGCCTCCCCAGGAA  
GGCTGCAGGCACTGCGGATCGTGCGCCCTCACATGATTATCCTGAGGCCCTTCTGCAG  
ATGCCATCAGGGCAGCACTCTGATGAGGTATTAGGGCACAGCACAGGGCTAAGCCAC  
CCTGTACTGGGCCAAGCGCTACAGGCAAAAAGGACACCACCGACGGGCATTTCAATTCATC  
[G,A]  
CTTTTATTTTTATATATTTTTGAGAGGGAGCCTCACTCTGTGCCCCAGGCTGGAGTGCAG  
TGGCGCATCTTGGCTCACTGCAACTTCTCCCTCCTGGGTTT

[illegible]